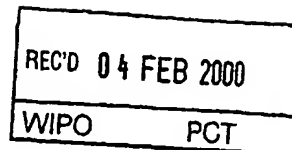


FR 00 / 144



EJV



# BREVET D'INVENTION

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

**PRIORITY  
DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

**COPIE OFFICIELLE**

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le 26 JAN. 2000

Pour le Directeur général de l'Institut  
national de la propriété industrielle  
Le Chef du Département des brevets

Martine PLANCHE

INSTITUT  
NATIONAL DE  
LA PROPRIÉTÉ  
INDUSTRIELLE

SIEGE  
26 bis, rue de Saint Petersburg  
75800 PARIS Cedex 08  
Telephone 01 53 04 53 04  
Telecopie 01 42 93 99 30



26 bis, rue de Saint Petersburg  
75800 Paris Cedex 08  
Téléphone : 01 53 04 53 04 Télécopie : 01 42 93 59 30

# BREVET D'INVENTION, CERTIFICAT D'UTILITÉ

Code de la propriété intellectuelle-Livre VI



N° 55-1328

## REQUÊTE EN DÉLIVRANCE

Confirmation d'un dépôt par télécopie ☒

Cet imprimé est à remplir à l'encre noire en lettres capitales

<b>DATE DE REMISE DES PIÈCES</b> 21 JAN. 1999 <b>N° D'ENREGISTREMENT NATIONAL</b> 99 00888 <b>DÉPARTEMENT DE DÉPÔT</b> 4 <b>DATE DE DÉPÔT</b> 21 JAN. 1999		<b>1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE</b>  CABINET GERMAIN & MAUREAU  B.P. 6153  69466 LYON CEDEX 06  n° du pouvoir permanent : MD/MK/B05B3294    références du correspondant : 04 72 69 84 30    téléphone :	
<b>2 DEMANDE</b> Nature du titre de propriété industrielle <input checked="" type="checkbox"/> brevet d'invention <input type="checkbox"/> demande divisionnaire <input type="checkbox"/> certificat d'utilité <input type="checkbox"/> transformation d'une demande de brevet européen <input type="checkbox"/> brevet d'invention <input type="checkbox"/> certificat d'utilité n° Établissement du rapport de recherche <input type="checkbox"/> différé <input checked="" type="checkbox"/> immédiat Le demandeur, personne physique, requiert le paiement échelonné de la redevance <input type="checkbox"/> oui <input type="checkbox"/> non Titre de l'invention (200 caractères maximum)  FRAGMENT NUCLEIQUE ENDOGENE ASSOCIE A UNE MALADIE AUTO-IMMUNE, PROCEDE DE MARQUAGE ET REACTIF			
<b>3 DEMANDEUR (S)</b> n° SIREN : 6.7.3.6.2.0.3.9.9    code APE-NAF : Nom et prénoms (souligner le nom patronymique) ou dénomination  BIO MERIEUX  Nationalité (s) Française Adresse (s) complète (s)  Chemin de l'Orme 69280 MARCY L'ETOILE  Pays FRANCE		Forme juridique  S.A.	
En cas d'insuffisance de place, poursuivre sur papier libre <input type="checkbox"/>			
<b>4 INVENTEUR (S)</b> Les inventeurs sont les demandeurs <input type="checkbox"/> oui <input checked="" type="checkbox"/> non    Si la réponse est non, fournir une désignation séparée			
<b>5 RÉDUCTION DU TAUX DES REDEVANCES</b> <input type="checkbox"/> requise pour la 1ère fois <input type="checkbox"/> requise antérieurement au dépôt : joindre copie de la décision d'admission			
<b>6 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE</b> pays d'origine    numéro    date de dépôt    nature de la demande			
<b>7 DIVISIONS</b> antérieures à la présente demande    n°    date    n°    date			
<b>8 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE</b> (nom et qualité du signataire)  Mireille DIDIER CPI 971202		<b>SIGNATURE DU PRÉPOSE À LA RÉCEPTION</b> <b>SIGNATURE APRES ENREGISTREMENT DE LA DEMANDE À L'INPI</b>  D. GIRAUD	

La loi n°78-17 du 6 janvier 1978 relative à l'information des citoyens et aux libertés s'applique aux requêtes faites à ce formulaire. Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'INPI.

# DOCUMENT COMPORTANT DES MODIFICATIONS

PAGE(S) DE LA DESCRIPTION OU DES REVENDECATIONS OU PLANCHE(S) DE DESSIN			R.M.*	DATE DE LA CORRESPONDANCE	TAMPON DATEUR DU CORRECTEUR
Modifiée(s)	Supprimée(s)	Ajoutée(s)			
phg			X	29/7/98	J P M - 04 AOUT 1998

Un changement apporte a la redaction des revendications d'origine, sauf si celui-ci decoule des dispositions de l'article R.612-36 du code de la Propriete Intellectuelle, est signale par la mention «R.M.» (revendications modifiees).

La présente invention se rapporte à un fragment nucléique endogène de type rétroviral intégré à l'ADN du génome humain.

Les rétrovirus sont des virus à ARN qui se répliquent par un processus dit de transcription inverse, médié par une ADN polymérase ARN dépendante dénommée reverse transcriptase (RT), codée par le gène *pol*. L'ARN rétroviral comprend également au moins deux gènes additionnels qui sont les gènes *gag* et *env*. Le gène *gag* code pour les protéines du squelette, c'est à dire pour la matrice, la capside et la nucléocapside. Le gène *env* code pour les protéines d'enveloppe. La transcription est régulée par des régions promotrices situées au niveau des LTR (Long Terminal Repeat) encadrant les extrémités 5' et 3' terminales du génome rétroviral.

Au cours de l'évolution les humains ou leurs ancêtres ont intégré dans leur génome du matériel d'origine rétrovirale suite à une infection. En effet, lors de l'infection d'une cellule, la reverse transcriptase fait une copie ADN de l'ARN rétroviral, cette copie d'ADN pouvant alors éventuellement s'intégrer dans le génome humain. Les rétrovirus peuvent infecter les cellules germinales et se transmettre ainsi aux générations futures par transmission Mendélienne verticale. On parle alors de rétrovirus endogènes qui sont présents sous la forme d'ADN proviral intégré dans le génome de toutes les cellules humaines. La plupart des rétrovirus endogènes sont silencieux ou défectifs. Toutefois certains d'entre eux ont pu conserver tout ou partie de leurs propriétés initiales et peuvent être activés dans des conditions particulières. L'expression des rétrovirus endogènes peut aller de la transcription de gènes viraux jusqu'à la production de particules virales.

Ces rétrovirus endogènes peuvent être associés directement ou indirectement au développement de certaines pathologies.

Des structures rétrovirales endogènes peuvent se retrouver sous une forme complète LTR-gag-pol-env-LTR ou sous des formes tronquées.

Ainsi, dans une précédente demande de brevet  
 5 (PCT/FR98/01442), la Demanderesse a criblé une banque d'ADNc à l'aide d'une sonde Ppol-MSRV (SEQ ID NO:18) et détecté des clones chevauchants qui lui ont permis de reconstruire un ARN génomique putatif de 7582 nucléotides. Cet ARN génomique présente une structure R-U5-gag-pol-env-  
 10 U3-R. Une interrogation « blastn » sur plusieurs bases de données à l'aide du génome reconstruit a permis de montrer qu'il existe une quantité importante de séquences génomiques (ADN) apparentées dans le génome humain qui sont trouvées sur plusieurs chromosomes. Ainsi, la  
 15 Demanderesse a mis en évidence l'existence de structures partielles de type rétroviral dans le génome humain et envisagé leur rôle potentiel dans le développement de maladies auto-immunes, dans des insuccès de grossesse ou des états de grossesse pathologiques.

20 A titre d'exemple on peut citer comme maladie auto-immune la sclérose en plaques, la polyarthrite rhumatoïde, le lupus érythémateux disséminé, le diabète insulino-dépendant et/ou les pathologies qui leur sont associées.

25 L'isolement et le séquençage de fragments d'ADNc chevauchants et l'identification de clones génomiques (ADN) correspondant aux clones d'ADN isolés, décrits dans la demande de brevet PCT précitée de la Demanderesse sont incorporés ici à titre de référence.

30 Isolement et séquençage de fragments d'ADNc chevauchants :

Les informations concernant l'organisation de la nouvelle famille de rétrovirus endogènes dénommée par la Demanderesse HERV-W ont été obtenues en testant une banque  
 35 d'ADNc placentaire (Clontech cat#HL5014a) avec les sondes Ppol-MSRV (SEQ ID NO:18) et Penv-C15 (SEQ ID NO:19) et en

pratiquant ensuite une technique de "gene walking" à l'aide des nouvelles séquences obtenues. Les expériences ont été mises en oeuvre en se référant aux préconisations du fournisseur de la banque. Des amplifications PCR sur  
 5 ADN ont également été exploitées pour comprendre cette organisation.

Les clones suivants ont été sélectionnés et séquencés :

- Clone cl.6A2 (SEQ ID NO:20) : région 5' non  
 10 traduite de HERV-W et une partie de gag
- Clone cl.6A1 (SEQ ID NO:21): gag et une partie  
 de pol
- Clone cl.7A16 (SEQ ID NO:22): Région 3' de pol
- Clone cl.Pi22 (SEQ ID NO:23): région 3' de pol  
 15 et début de env
- Clone cl.24.4 (SEQ ID NO:24) : ARN épissé  
 comprenant une partie de la région 5' non traduite de  
 HERV-W, la fin de pol et la région 5' de env
- Clone cl.C4C5. (SEQ ID NO:25) : fin de env et  
 20 région 3' non traduite de HERV-W
- Clone cl.PH74 (SEQ ID NO:26) : ARN sous-  
 génomique : région 5' non traduite de HERV-W, fin de pol,  
 env, et région 3' non traduite de HERV-W
- Clone cl.PH7 (SEQ ID NO:27) : ARN multi-épissé  
 25 : région 5' non traduite de HERV-W, fin de env et région  
 3' non traduite de HERV-W.
- Clone cl.Pi5T (SEQ ID NO:28) : gène pol partiel  
 et région U3-R
- Clone cl.44.4 (SEQ ID NO:29) : région R-U5,  
 30 gène gag et gène pol partiel.

A l'aide de ces clones, en procédant à des alignements de séquences, un modèle de séquence totale de HERV-W a été élaboré. Les ARN épissés ont été mis en évidence ainsi que les sites potentiels donneurs et  
 35 accepteurs d'épissage. Par étude de similitude avec des

rétrovirus existants, les entités LTR, gag, pol et env ont été définies.

L'organisation génétique putative de HERV-W sous forme ARN est la suivante (SEQ ID NO:30):

5	gène	1..7582
	Localisation des clones sur la séquence ARN génomique reconstruite	
		cl.6A2 (1321 pb) 1-1325 ;
10		cl.PH74 (535+2229= 2764 pb) 72-606 et 5353-7582;
		cl.24.4 (491+1457= 1948 pb); 115-606 et 5353-6810;
		cl.44.4 (2372 pb) 115-2496 ;
		cl.PH7 (369+297= 666pb) 237-606 et 7017-7313;
15		cl.6A1 (2938 pb) 586-3559.;
		cl.Pi5T (2785+566= 3351 pb) 2747-5557 et 7017-7582;
		cl.7A16 (1422 pb) 2908-4337;
20		cl.Pi22 (317+1689 = 2006 pb) 3957-4273 et 4476-6168;
		cl.C4C5 (1116 pb) 6467-7582
	5'LTR	1..120
		/note="R of 5'LTR (extrémité 5' incertaine"
25		121..575
		/note="U5 of 5'LTR"
	divers	579..596
		/note="PBS primer binding site pour tRNA-W"
30	divers	606
		/note="jonction d'épissage (site donneur d'épissage ATCCAAAGTG-GTGAGTAATA et site accepteur d'épissage CTTTTTTCAG-ATGGGAAACG clone RG083M05, GenBank accession
35		AC000064) "
	divers	5353

```

                                /note="site accepteur d'épissage pour
                                l'ORF1 (env)"
divers                          5560
                                /note="site donneur d'épissage"
5  ORF                          5581..7194
                                /note="ORF1 env 538 AA"
                                /produit=="enveloppe"
divers                          7017
                                /note="site accepteur d'épissage pour ORF2
10 ORF                          7039..7194
                                /note="ORF2 52 AA"
ORF                             7112..7255
                                /note="ORF3 48 AA"
15 divers                       7244..7254
                                /note="PPT polypurine tract"
3'LTR                           7256..7582
                                /note=="U3-R of 3' LTR (jonction U3-R
                                indéterminée)
20 divers                       7563..7569
                                signal de polyadénylation

```

Identification de clones génomiques (ADN)  
correspondant aux clones d'ADN isolés :

25 Une interrogation "blastn" sur plusieurs bases de  
données, à l'aide du génome reconstruit, a montré qu'il  
existe une quantité importante de séquences apparentées  
dans le génome humain. Environ 400 séquences ont été  
identifiées dans GenBank et plus de 200 séquences dans la  
30 banque EST, la plupart en anti-sens. Les 4 séquences les  
plus significatives en taille et en similitude sont les  
séquences des clones génomiques (ADN) suivants :

le clone humain RG083M05 (gb AC000064) dont la  
localisation chromosomique est 7q21-7q22,

le clone humain BAC378 (gb U85196, gb AE000660)  
correspondant au locus alpha delta du récepteur des  
cellules T, localisé en 14q11-12,

le cosmide humain Q11M15 (gb AF045450)  
5 correspondant à la région 21q22.3 du chromosome 21,  
le cosmide U134E6 (embl Z83850) sur le chromosome  
Xq22.

La localisation des régions alignées pour chacun  
des clones est indiquée et l'appartenance à un chromosome  
10 est indiquée entre crochets (figure 6). Le pourcentage de  
similitude (sans les larges délétions) entre les 4  
séquences et l'ARN génomique reconstruit est indiqué,  
ainsi que la présence de séquences répétées à chaque  
extrémité du génome et la taille des plus grandes trames  
15 de lecture (ORF). Des séquences répétées ont été trouvées  
aux extrémités de 3 de ces clones. La séquence  
reconstruite est contenue intégralement à l'intérieur du  
clone RG083M05 (9,6 Kb) et présente une similitude de 96%.  
Cependant le clone RG083M05 présente une insertion de 2 Kb  
20 située immédiatement en aval de la région 5' non traduite  
(5' UTR). Cette insertion est également trouvée dans deux  
autres clones génomiques qui présentent une délétion de  
2,3 Kb immédiatement en amont de la région 3' non traduite  
(3' UTR). Aucun clone ne contenait les trois trames de  
25 lecture (ORFs) gag, pol et env fonctionnelles. Le clone  
RG083M05 montre une ORF de 538 acides aminés (AA)  
correspondant à une enveloppe entière. Le cosmide Q11M15  
contient deux grandes ORFs contigües de 413 AA (trame 0)  
et 305 AA (trame +1) correspondant à une polyprotéine pol  
30 tronquée.

On a maintenant trouvé et isolé un fragment  
nucléique endogène, intégré à l'ADN du génome humain, qui  
comprend ou consiste en au moins une partie du gène gag  
d'un rétrovirus endogène associé à une maladie auto-  
35 immune, ou à des insuccès de grossesse ou pathologies de  
la grossesse, cette partie au moins codant, directement ou

indirectement, pour un produit d'expression. Bien entendu, l'invention comprend également le complémentaire dudit fragment.

Avantageusement, le fragment précédemment défini  
5 répond en outre à au moins l'une quelconque des caractéristiques suivantes :

il comprend, ou consiste en, ledit gène gag entier ;

ladite partie du fragment au moins code pour la  
10 matrice et la capsid ;

il comprend, ou consiste en, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, ou le complémentaire de l'une quelconque desdites séquences ;

il est localisé sur au moins l'un des chromosomes  
15 humain 1, 3, 6, 7 et 16, de préférence il est localisé au moins sur le chromosome 3 ;

le produit d'expression de ladite partie est de l'ARN messager ;

le produit d'expression de ladite partie est  
20 immunologiquement reconnu par des anticorps présents dans un échantillon biologique d'un patient atteint d'une maladie auto-immune, telle que la sclérose en plaques ; de préférence, le fluide biologique est choisi parmi le sérum, le plasma, le liquide synovial et l'urine.

25 Un autre objet de l'invention est un produit de transcription endogène, à l'état isolé, susceptible d'être obtenu par transcription d'au moins ladite partie du gène gag d'un fragment de l'invention.

L'invention concerne aussi un procédé de  
30 détection de séquences nucléotidiques endogènes appartenant à un fragment de l'invention comprenant les étapes suivantes :

on effectue au préalable une étape d'extraction de l'ADN cellulaire d'un tissu ou fluide biologique, on  
35 réalise au moins un cycle d'amplification de l'ADN

cellulaire, telle que par PCR, au moyen d'amorces notamment choisies parmi SEQ ID NO:4 à SEQ ID NO:9 et SEQ ID NO :12 à SEQ ID NO :17,

on met en contact l'ADN cellulaire présent dans  
5 l'échantillon avec une sonde déterminée susceptible de s'hybrider avec un fragment tel que défini précédemment et de former un complexe d'hybridation, ladite sonde comprenant au moins 15 nucléotides, de préférence 17 et avantageusement 19 nucléotides contigus de SEQ ID NO :3 ou  
10 consistant en SEQ ID NO :3, dans des conditions appropriées permettant l'hybridation, en particulier dans des conditions de forte stringence, et

on détecte les complexes d'hybridation formés par tout moyen approprié.

15 Avantageusement, la sonde est marquée par un traceur, tel que par exemple un traceur radioactif ou une enzyme.

L'invention concerne aussi un procédé de détection de séquences nucléotidiques endogènes  
20 appartenant à un fragment de l'invention comprenant les étapes suivantes :

on effectue au préalable une étape d'extraction de l'ADN cellulaire d'un tissu ou fluide biologique, on réalise au moins un cycle d'amplification de l'ADN  
25 cellulaire, telle que par PCR, au moyen d'amorces notamment choisies parmi SEQ ID NO:4 à SEQ ID NO:9 et SEQ ID NO :12 à SEQ ID NO :17,

on effectue une étape de transcription / traduction in vitro du produit amplifié, et  
30 on fait réagir le produit issu de l'étape de transcription / traduction avec un sérum ou plasma d'un patient présentant une maladie auto-immune.

L'invention concerne encore un procédé pour l'étude et/ou le suivi d'une prolifération cellulaire des  
35 cellules T in vitro selon lequel on met en contact les cellules T d'un patient avec soit des produits de

transcription / traduction (SEQ ID NO :31), tels qu'obtenus selon le procédé ci-dessus, soit des peptides de synthèse dérivés de ou appartenant à SEQ ID NO :31.

Un autre objet de l'invention est un procédé de  
5 marquage moléculaire in situ de chromosomes isolés de patients dans lequel on utilise une sonde marquée, par tout traceur approprié, comprenant tout ou partie de SEQ ID NO :3.

L'invention concerne aussi :

10 Une protéine recombinante obtenue à partir d'une cassette d'expression dans un hôte bactérien caractérisée en ce que sa séquence protéique consiste en SEQ ID NO :31 ; en particulier l'hôte bactérien est *E. coli*.

Un réactif pour la détection d'une maladie auto-  
15 immune ou un suivi de grossesse comprenant au moins un fragment ou une protéine de l'invention ;

L'utilisation d'un fragment ou d'une protéine de l'invention pour détecter, dans un échantillon biologique, une susceptibilité à une maladie auto-immune ou un suivi  
20 de grossesse ; en particulier, la maladie auto-immune est la sclérose en plaques.

Avant d'exposer la présente invention plus en détail, la définition de certains termes employés dans la description et les revendications est donnée.

25 L'expression « produit d'expression » signifie tout produit dérivé de l'ADN rétroviral intégré dans le génome humain incluant les produits de la transcription (ARN messenger) et les produits issus de la traduction de l'ARN messenger obtenu. Dans ce dernier cas, et à titre  
30 d'exemple, le produit peut être un peptide, une protéine fonctionnelle ou fonctionnalisable, c'est-à-dire susceptible de devenir fonctionnelle.

Par partie codant, directement ou indirectement, pour un produit d'expression, on entend une partie qui à  
35 elle seule comprend au moins tout ou partie d'un cadre de lecture ouvert duquel on peut déduire une séquence

d'acides aminés, et dont la capacité codante peut être induite par des éléments tels que par exemple ceux susceptibles d'avoir une activité promotrice. Cette définition inclut la variabilité qui peut être retrouvée  
 5 dans la séquence nucléique codante, sous réserve que les conditions ci-dessus soient respectées.

**Exemple 1: Localisation du gène *gag* de la famille  
 HERV-W sur les chromosomes humains par la technique de  
 10 Southern blot**

Afin de localiser le gène *gag* de la famille HERV-W, une sonde correspondante à ce gène de MSRV a été hybridée sur une membrane de Nylon (Hybond® N+, Amersham) contenant 5 µg d'ADN de 24 hybrides somatique de cellules  
 15 [humain x rongeurs] (de l'ADN génomique humain isolé: 22 chromosomes autosomes et 2 chromosomes sexuels) et 3 ADN contrôles (humain, souris et hamster) digérés par l'enzyme de restriction EcoRI.

La sonde utilisée est la suivante : Pgag-C12  
 20 identifiée par SEQ ID NO:3 correspondant à la région codante (de 1056 pb) du clone MSRV gag C12.

1.1- Obtention du clone 2, C12, contenant en 3' une partie homologue au gène *pol*, correspondant au gène  
 25 protéase, et une partie homologue au gène *gag* correspondant à la nucléocapside et une région 5' codante, correspondant au gène *gag*, plus spécifiquement la matrice et la capside de MSRV-1.

Une amplification par PCR a été effectuée sur de  
 30 l'ARN total extrait à partir de 100 µl de plasma d'un patient atteint de SEP. Un témoin eau, traité dans les mêmes conditions, a été utilisé comme contrôle négatif. La synthèse de cDNA a été réalisée avec 300 pmoles d'une amorce aléatoire (Gibco-BRL, France) et la transcriptase  
 35 inverse « Expand RT » (Boehringer Mannheim, France) selon les conditions préconisées par la société. Une

amplification par PCR (polymerase chain reaction) a été effectuée avec l'enzyme Taq polymerase (Perkin Elmer, France) en utilisant 10 µl de cDNA dans les conditions suivantes : 94°C 2 min, 55°C 1 min, 72°C 2 min puis 94°C  
 5 1 min, 55°C 1 min, 72°C 2 min pendant 30 cycles et 72°C pendant 7 min et avec un volume réactionnel final de 50 µl.

Les amorces utilisées pour l'amplification par PCR sont les suivantes :

- 10 - amorce 5', identifiée par SEQ ID NO:4  
 5' CGG ACA TCC AAA GTG ATG GGA AAC G 3' ;
- amorce 3', identifiée par SEQ ID NO:5  
 5' GGA CAG GAA AGT AAG ACT GAG AAG GC 3'

Une deuxième amplification par PCR dite  
 15 « nichée » a été réalisée avec des amorces 3' et 5' situées à l'intérieur de la région déjà amplifiée. Cette deuxième PCR a été effectuée dans les mêmes conditions expérimentales que celles utilisées lors de la première PCR, en utilisant 10 µl du produit d'amplification issu de  
 20 la première PCR.

Les amorces utilisées pour l'amplification par PCR nichée sont les suivantes :

- amorce 5', identifiée par SEQ ID NO:6  
 5' CCT AGA ACG TAT TCT GGA GAA TTG GG 3' ;
- 25 - amorce 3', identifiée par SEQ ID NO:7  
 5' TGG CTC TCA ATG GTC AAA CAT ACC CG 3'

Un produit d'amplification de 1511 pb a été obtenu à partir de l'ARN extrait du plasma de patient SEP. Le fragment correspondant n'a pas été observé pour le  
 30 témoin eau. Ce produit d'amplification a été cloné de la façon suivante.

L'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning®. Les 2 µl de solution d'ADN ont été mélangés avec 5 µl d'eau distillée stérile, 1 µl d'un  
 35 tampon de ligation 10 fois concentré « 10X Ligation Buffer », 2 µl de « PCR® Vector » (25 ng/ml) et 1 µl de

« T4 DNA Ligase ». Ce mélange a été incubé une nuit à 14°C. Les étapes suivantes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). Le mélange de ligation a été étalé après transformation de la ligation dans des bactéries *E. coli*. A la fin de la procédure, les colonies blanches de bactéries recombinantes ont été repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de « minipréparation d'ADN » (J. Sambrook, E.F. Fritsch and T. Maniatis, *Molecular Cloning*, a laboratory manual, Cold Spring Harbour Laboratory Press, 1989). La préparation de plasmide de chaque colonie recombinante a été coupée par l'enzyme de restriction EcoRI et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage « Prism® Ready Reaction Amplitaq® FS, DyeDeoxy™ Terminator » (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone obtenu, dénommé Cl2, permet de définir une région de 1511 pb présentant une phase ouverte de lecture dans la région N-terminale de 1056 pb (SEQ ID NO:3) codante pour 359 acides aminés (SEQ ID NO:31) correspondant aux régions matrice et capsid du gène *gag*.

La séquence nucléotidique de Cl2 est identifiée par SEQ ID NO:1. Elle est représentée à la figure 2 avec les trames de lecture potentielles en aminoacides.

### 1.2- Obtention de la sonde gag cl2 MSRV

La sonde a été obtenue après amplification par PCR, à partir du plasmide pCR<sup>™</sup> vector (kit TA Cloning<sup>®</sup>, Invitrogen) contenant l'insert du clone: gag cl2 de MSRV, avec la Taq polymérase (Perkin Elmer, France) dans les conditions suivantes : 94°C 1 min, 55°C 1 min, 72°C 2 min pendant 35 cycles et 72°C pendant 7 min et avec un volume réactionnel final de 100 µl.

10 Les amorces utilisées pour l'amplification par PCR sont les suivantes :

- amorce 5', identifiée par SEQ ID NO:12

5'-CTA GAA CGT ATT CTG GAG AAT TGG GA-3'

- amorce 3', identifiée par SEQ ID NO:13

15 5'-CCT AAG GCA GAC TTT TGA AG -3'

Un produit d'amplification de 1056 pb a été obtenu pour gag cl2 de MSRV.

Après l'amplification par PCR, le fragment a été analysé en gel d'agarose 1%. Le fragment détecté sous lumière UV, après marquage du gel, au bromure d'éthidium, a été coupé et marqué au [ $\alpha$ -P<sup>32</sup>] à l'aide des amorces aléatoires (Gibco-BRL, France) conformément aux instructions du kit « Ready to go DNA labelling » (Pharmacia Biotech). Les nucléotides non incorporés ont été éliminés avec une colonne G-50 Quick Spin (Boehringer, Mannheim).

### 1.3- Southern blot

Les conditions d'hybridation sont les suivantes :

Après 4 heures de pré-hybridation (en 5x SSC, 1x Denhardt's, 0,1% SDS, 50% formamide, 20 mM Tris-HCl pH = 7,5, 0,1 mg/ml d'ADN de sperme de hareng), la membrane de Nylon contenant les chromosomes humains a été hybridée (en 5x SSC, 1x Denhardt's, 0,1% SDS, 50% formamide, 20 mM Tris-HCl pH = 7,5, 0,1 mg/ml d'ADN de sperme de hareng, 5% sulfate de dextran) pendant 18 heures à 42°C, avec la sonde d'ADN de 1056 pb (SEQ ID NO:3) de

gag c12 marquée au  $^{32}\text{P}$ . Après hybridation, la membrane (The BIOS Monochromosomal Somatic Cell Hybrid blot, de Quantum Bioprobe) hybridée avec la sonde gag a été lavée en solution 2x SSC, 0,2% SDS 2 fois 15 min, à température  
5 ambiante, et (en 0,2x SSC - 0,2% SDS) 2 fois 15 min, à 45°C. Après lavage, la membrane a été exposée au film aux rayons X, à -80°C en présence d'écran amplificateur.

Les résultats sont présentés dans le tableau 1 ci-après.

**Exemple 1** : Localisation chromosomique de la région gag HERV-W par Southern blot.

n°chromo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	sou- -ris	ham- -ster
rodent parent	m	h	h	h	h	h	h	h	h	h	h	h	h	h	h	m	m	h	h	m	m	h	h	h	0	0
sonde gag	5	0	6	6	5	3	2	3	2	4	3	6	3	1	3	0	3	2	1	0	4	0	4	0	0	0

mouse (m), hamster (h) : cellules receveuses de l'ADN de chromosomes humains.

Le chiffre indiqué en-dessous de chaque chromosome correspond au nombre de bandes rencontrées. Total de copies pour le gène gag = 66

Tableau 1

Exemple 2: Amplification par PCR du gène *gag* de la famille HERV-W sur chacun des chromosomes humains isolés ; vérification de la spécificité des amplifications par Southern blot ; test de transcription-traduction « in vitro » (PTT) à partir des produits de PCR, afin de vérifier la capacité codante, repérer quels sont les chromosomes humains présentant des trames de lecture ouvertes pour le gène *gag* de la famille HERV-W.

## 10 2.1- Amplification par PCR

Pour l'amplification du gène *gag* HERV-W, une PCR a été effectuée sur chaque chromosome humain isolé [NIGMS human/rodent somatic cell hybrid panel #2. The human monochromosomal NIGMS somatic hybrid mapping panel #2, décrit par H.L. Drwinga et al et B.L. Dubois et al] obtenu de Coriell Institute (Camden, NJ)] avec l'enzyme Taq polymérase (Perkin Elmer, France) en utilisant: 40 pmol de chaque amorce, 25 mM de chaque dNTP (Pharmacia), 2,5 mM MgCl<sub>2</sub>, 2,5 U de Taq polymérase dans le tampon PCR standard (Perkin Elmer), et 300 ng DNA de chromosome isolé dans un volume final de 100 µl. Les conditions de PCR pour amplifier la région *gag* sont les suivantes : 3 min à 94°C; puis 1 min à 94°C, 1 min à 55°C, 3 min à 72°C, pendant 30 cycles et 7 min à 72°C.

25 Les amorces utilisées pour l'amplification, par PCR, du gène *gag* à partir d'un ATG introduit dans la séquence *gag* HERV-W sur chaque chromosome humain isolé sont les suivantes :

- amorce 5', identifiée par SEQ ID NO:14  
30 5'-TTT GGT AAT ACG ACT CAC TAT AGG GCA GCC ACC ATG GGA AAC GTT CCC CCC GAG-3'

L'amorce contient la séquence du promoteur T7 RNA polymérase, un « spacer », la séquence Kozak (site d'initiation à la traduction chez les eucaryotes) et la séquence 5' *gag* à partir de l'ATG de HERV-W.

- amorce 3', identifiée par SEQ ID NO:15

5'-TTTTTTTTTTTTTTTTTTTTTTCAGGCTGCGCCAGTGTCCAGGAGAC-3'

L'amorce contient une queue de poly-A (pour stabiliser la transcription de l'ARN, représentée par 18 bases T), un codon stop (représenté par TCA) et la  
 5 séquence du gène protéase (G+E+A) de MSRV-1.

Pour l'amplification du gène *gag* HERV-W en utilisant des oligonucléotides définis dans les régions LTR et protéase de HERV-W avec l'enzyme *Taq* polymérase (Perkin Elmer, France), les conditions de PCR ont été les  
 10 suivantes : 3 min à 94°C; puis 1 min à 94°C, 1 min à 60°C, 2 min à 72°C, 35 cycles ; suivi de 7 min à 72°C, avec 50 ng de chaque ADN monochromosomal.

Les amorces utilisées pour l'amplification, par PCR, du gène *gag* à l'aide de l'oligonucléotide défini dans  
 15 la séquence LTR HERV-W sur chaque chromosome humain isolé, sont les suivantes :

- amorce 5', identifiée par SEQ ID NO:16

5'-TGTCGCTGTGCTCCTGATC-3'

- amorce 3', identifiée par SEQ ID NO:17

20 5'-TTTTTTTTTTTTTTTTTTTTTTCAGGCTGCGCCAGTGTCCAGGAGAC-3'

L'amorce contient une queue de poly-A (pour stabiliser la transcription de l'ARN, représenté par 18 bases T), un codon stop (représenté par TCA) et la séquence du gène protéase de G+E+A de MSRV-1.

25 Les amplification par PCR ont été effectuées dans l'appareil MJ Research PTC200, Peltier Thermal cycler. Les produits de PCR (10 µl de chaque produit PCR) ont été analysés dans un gel 1% agarose en TBE 1X(Tris-HCl, borate, EDTA). Afin de vérifier la spécificité des  
 30 produits d'amplification, 3 µl de chaque produit PCR ont été analysés en gel d'agarose et puis transférés sur une membrane de Nylon (Hybond®-N<sup>+</sup>, Amersham) (Southern blot) à l'aide de NaOH 0,4N. L'hybridation avec la sonde *gag* cl2 (1056 pb) (J. Sambrook et al, 1989) a été effectuée dans  
 35 les conditions suivantes : après 4 heures de pré-hybridation (en 5x SSC, 1X Denhardt's, 0,1% SDS, 50%

formamide, 20 mM Tris-HCl pH = 7,5, 0,1 mg/ml d'ADN de sperme de hareng), la membrane de Nylon a été hybridée (en 5x SSC, 1x Denhardt's, 0,1% SDS, 50% formamide, 20 mM Tris-HCl pH = 7,5, 0,1 mg/ml d'ADN de sperme de hareng, 5% sulfate de dextran) pendant 18 heures à 42°C, avec la sonde d'ADN gag marquée au <sup>32</sup>P. Les produits de PCR gag de chaque chromosome humain isolé ont été lavés en solution de 2x SSC, 0,2% SDS 1 fois, 15 min à température ambiante; 0,2x SSC, 0,1% SDS 2 fois 15 min chaque lavage à 65°C, en 0,1x SSC, 0,1% SDS 2 fois 15 min chaque à 65°C, et en 0,1x SSC, 0,1% SDS 2 fois 30 min chaque à température ambiante.

Une partie du volume restant (4 µl) des produits d'amplification par PCR a été utilisée pour le test de transcription-traduction « in vitro » PTT, (Roest PAM et al, 1993) (Promega, France). Le volume restant a été utilisé pour le clonage dans le vecteur pCR<sup>®</sup> 2.1-TOPO (Invitrogen) conformément aux instructions du kit, et pour le séquençage avec la méthode préconisée pour l'utilisation du kit de séquençage « PRISM<sup>™</sup> Ready Reaction Amplitaq<sup>®</sup> FS, DyeDeoxy<sup>™</sup> Terminator » (Applied Biosystems, réf. 402119) et le séquençage automatique a été réalisé sur les appareils 373A et 377 Applied Biosystems, selon les instructions du fabricant.

La partie codée (SEQ ID NO:31) par le fragment de 2009 pb (SEQ ID NO:2) a été amplifié par PCR avec l'enzyme *Pwo* (5U/µl) (Boehringer Mannheim, France) en utilisant 1 µl de la mini-préparation de l'ADN du clone gag (SEQ ID NO :3) sous les conditions suivantes : 95°C 1 min, 60°C 1 min, 72°C 2 min pendant 25 cycles et avec un volume réactionnel final de 50 µl à l'aide des amorces:

- amorce 5' (*Bam* HI) (SEQ ID NO :8):  
5' ATG GGA AAC GTT CCC CCC GAG 3' (21 mers), et
- amorce 3' (*Hind* III), identifiée par SEQ ID NO:9  
5 GGC CTA AGG CAG ACT TTT GAA 3' (21 mers)

Le fragment obtenu après PCR, a été linéarisé par Bam HI et Hind III et sous-cloné dans les vecteurs d'expression pET28C et pET21C (NOVAGEN) linéarisé par Bam HI et Hind III. Le séquençage de l'ADN du fragment de 1089 pb dans les deux vecteurs d'expression a été réalisé selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISM<sup>TM</sup> Ready Reaction Amplitaq® FS, DyeDeoxy<sup>TM</sup> Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

L'expression de la séquence nucléotidique du fragment de 1089 pb du clone gag par les vecteurs d'expression pET28C et pET21C sont identifiées par respectivement SEQ ID NO:10 et SEQ ID NO:11.

## 2.2- Test de transcription-traduction « in vitro » (PTT, Promega)

Ce test a été effectué pour repérer les chromosomes humains présentant des trames de lecture ouvertes pour le gène gag de la famille HERV-W.

Dans un volume réactionnel de 25 µl, un mélange contenant 12,5 µl de lysat de réticulocyte de lapin TNT® (Promega), 1 µl de tampon de réaction TNT® (Promega), 0,5 µl d'ARN polymérase TNT® (Promega), 0,5 µl de mélange à 1mM des acides aminés moins méthionine, 2 µl de <sup>35</sup>S-méthionine (1.000 Ci/mmol) à 10 mCi/µl (Amersham), 0,5 µl d'inhibiteur de ribonucléase RNasin® à 40 U/µl, 4 µl de produits d'amplification par PCR (équivalent à 1 µg) de chaque chromosome humain et 4 µl d'eau. Ce mélange a été incubé à 30°C pendant 90 min.

Les protéines gag correspondant aux produits de transcription/traduction du gène gag de la famille HERV-W de chaque chromosome humain, amplifié par PCR, ont été mises en évidence par électrophorèse sur gel de

polyacrylamide à 10% en présence de Dodécyl Sulfate de Sodium (SDS)-PAGE après exposition du gel au film aux rayons X à température ambiante en présence d'écran amplificateur.

- 5 Les résultats sont présentés dans le tableau 2 ci-après.

**Exemple 2 :** Recherche de la capacité codante de gène gag par PCR puis PTT "transcription et traduction in vitro" (protein trunkation test).

n°chromo rodent parent	n°chromo rodent parent	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	eau
PCR gag	gag	28 23 18		45 25 20 17	-		23 18 17	22					-	-		-	14					-		-		-

Le chiffre indiqué en-dessous de chaque chromosome correspond à la masse moléculaire (kDa) des protéines visualisés en gel de polyacrylamide en présence de SDS.

Tableau 2

**Exemple 3 : expression du clone gag chez *Escherichia coli*, et réaction avec des sérums humains**

On a exprimé la partie codante SEQ ID NO:2 chez *Escherichia coli* puis on a testé les produits ainsi  
 5 exprimés vis à vis de sérum de patients atteints de SEP, ainsi que de sérum de patients sains.

Les constructions pET28c-clone gag (1089 pb) et pET21C-clone gag (1089 pb) synthétisent, dans la souche bactérienne BL21 (DE3), une protéine en fusion N- et C-terminale pour le vecteur pET28C et C-Terminale pour le  
 10 vecteur pET21C avec 6 résidus histidine, de masse moléculaire apparente d'environ 45 kDa, mise en évidence par électrophorèse sur gel de polyacrylamide SDS-PAGE (U.K. Laemmli, Cleavage of structural proteins during the  
 15 assembly of the head of bacteriophage T4, Nature, 1970, 227 :680-685).

La réactivité de la protéine a été mise en évidence vis à vis d'un anticorps monoclonal anti-Histidine (DIANOVA) par la technique de Western blot  
 20 (H. Towbin et al., Electrophoretic transfer of proteins from polyacrylamid gels to nitrocellulose sheets : procedure and some applications, Proc. Natl. Acad. Sci. USA., 1979, 76 :4350-4354).

Les protéines recombinantes pET28c-clone gag  
 25 (1089 pb) et pET21C-clone gag (1089 pb) ont été visualisées en SDS-PAGE dans la fraction insoluble après digestion enzymatique des extraits bactériens avec 50 µl de lysozyme (10 mg/ml) et lyse par ultrasons.

Les propriétés antigéniques des antigènes  
 30 recombinants pET28C-clone gag (1089 pb) et pET21C-clone gag (1089 pb) ont été testées par Western Blot après solubilisation du culot bactérien avec 2% SDS et 50 mM β-mercaptoéthanol. Après incubation avec les sérums de patients atteints de sclérose en plaques, les sérums des  
 35 témoins neurologiques et les sérums de témoins de centre de transfusion sanguine (CTS), les immunocomplexes ont été

détectés à l'aide d'un sérum de chèvre anti-IgG et anti-IgM humaines, couplé à la phosphatase alcaline.

Les résultats sont présentés dans le tableau 3 ci-après.

5

Tableau 3

Réactivité de sérums de patients atteints de sclérose en plaques et témoins avec la protéine recombinante de gag produite chez *E. coli*<sup>a</sup>

10	MALADIE	NOMBRE D'INDIVIDUS TESTÉS	NOMBRE D'INDIVIDUS POSITIFS
	SEP	15	6 2(+++), 2(++), 2(+)
	TEMOINS		
15	NEUROLOGIQUES	2	1(+++)
	TEMOINS		
	SAINS (CTS)	22	1(+/-)

(a) Les bandelettes contenant 1,5 µg d'antigène recombinant gag présentent une réactivité contre de sérums dilués au 1/100. L'interprétation de Western Blot est basée sur la présence ou absence d'une bande gag spécifique sur les bandelettes. Des contrôles positifs et négatifs sont inclus dans chaque expérience.

25

Ces résultats montrent que, dans les conditions techniques utilisées, environ 40% des sérums humains atteints de sclérose en plaques testés réagissent avec la protéine recombinante gag.

## LISTE DE SÉQUENCES

## (1) INFORMATIONS GENERALES:

## (i) DEPOSANT:

- 5 (A) NOM: BIOMERIEUX  
 (B) RUE: AUCUNE  
 (C) VILLE: MARCY L'ETOILE  
 (E) PAYS: FRANCE  
 (F) CODE POSTAL: 69280

- 10 (ii) TITRE DE L'INVENTION: Fragment nucléotidique endogène associé  
 à une maladie auto-immune, procédé de marquage et réactif.

(iii) NOMBRE DE SÉQUENCES: 31

## (iv) FORME DECHIFFRABLE PAR ORDINATEUR:

- (A) TYPE DE SUPPORT: Floppy disk  
 15 (B) ORDINATEUR: IBM PC compatible  
 (C) SYSTEME D'EXPLOITATION: PC-DOS/MS-DOS  
 (D) LOGICIEL: PatentIn Release #1.0, Version #1.30 (OEB)

## (2) INFORMATIONS POUR LA SEQ ID NO:1:

## 20 (i) CARACTERISTIQUES DE LA SÉQUENCE:

- (A) LONGUEUR: 1511 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## 25 (ii) TYPE DE MOLECULE: ADN

## (x1) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:1:

CCTAGAACGT ATTCTGGAGA ATTGGGACCA ATGTGACACT CAGACGCTAA GAAAGAAACG 60  
 ATTTATATTC TTCTGCAGTA CCGCCTGGCC ACAATATCCT CTTCAAGGGA GAGAAACCTG 120  
 GCTTCCTGAG GGAAGTATAA ATTATAACAT CATCTTACAG CTAGACCTCT TCTGTAGAAA 180  
 30 GGAGGGCAAA TGGAGTGAAG TGCCATATGT GCAAACCTTC TTTTCATTAA GAGACAACCTC 240  
 ACAATTATGT AAAAAAGTGTG GTTTATGCCC TACAGGAAGC CCTCAGAGTC CACCTCCCTA 300  
 CCCCAGCGTC CCCTCCCCGA CTCCTTCCTC AACTAATAAG GACCCCCCTT TAACCCAAAC 360  
 GGTCCAAAAG GAGATAGACA AAGGGGTAAA CAATGAACCA AAGAGTGCCA ATATTCCTCCG 420  
 ATTATGCCCC CTCCAAGCAG TGAGAGGAGG AGAATTCGGC CCAGCCAGAG TGCCTGTACC 480  
 35 TTTTCTCTC TCAGACTTAA AGCAAATTAA AATAGACCTA GGTAAATTCT CAGATAACCC 540  
 TGACGGCTAT ATTGATGTTT TACAAGGGTT AGGACAATCC TTTGATCTGA CATGGAGAGA 600

TATAATGTTA CTAATAATC AGACACTAAC CCCAATGAG AGAAGTGCCG CTGTAAGTGC 660  
 AGCCCCGAGAG TTTGGCGATC TTTGGTATCT CAGTCAGGCC AACAAATAGGA TGACAACAGA 720  
 GGAAAGAACA ACTCCCACAG GCCAGCAGGC AGTTCCAGT GTAGACCCTC ATTGGGACAC 780  
 AGAATCAGAA CATGGAGATT GGTGCCACAA ACATTTGCTA ACTTGCCTGC TAGAAGGACT 840  
 5 GAGGAAAAC AGGAAGAAGC CTATGAATTA CTCAATGATG TCCACTATAA CACAGGGAAA 900  
 GGAAGAAAAT CTTACTGCTT TTCTGGACAG ACTAAGGGAG GCATTGAGGA AGCATACCTC 960  
 CCTGTACCT GACTCTATTG AAGGCCAACT AATCTTAAAG GATAAGTTTA TCACTCAGTC 1020  
 AGCTGCAGAC ATTAGAAAAA ACTTCAAAAG TCTGCCTTAG GCCCGGAGCA GAACTTAGAA 1080  
 ACCCTATTTA ACTTGGCATC CTCAGTTTTT TATAATAGAG ATCAGGAGGA GCAGGCGAAA 1140  
 10 CGGGACAAAC GGGATAAAAA AAAAAGGGGG GGTCCACTAC TTTAGTCATG GCCCTCAGGC 1200  
 AAGCAGACTT TGGAGGCTCT GCAAAAGGGA AAAGCTGGGC AAATCAAATG CCTAATAGGG 1260  
 CTGGCTTCCA GTGCGGTCTA CAAGGACACT TTAATAAAGA TTATCCAAGT AGAAATAAGC 1320  
 CGCCCCCTTG TCCATGCCCC TTACGTCAAG GGAATCACTG GAAGGCCAC TGCCCCAGGG 1380  
 GATGAAGATA CTCTGAGTCA GAAGCCATTA ACCAGATGAT CCAGCAGCAG GACTGAGGGT 1440  
 15 GCCCCGGGCG AGCGCCAGCC CATGCCATCA CCCTCACAGA GCCCCGGGTA TGTTCGACCA 1500  
 TTGAGAGCCA A 1511

## (2) INFORMATIONS POUR LA SEQ ID NO:2:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- 20 (A) LONGUEUR: 2009 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN

## 25 (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:2:

ATACGACTAC TATAGGGCGA ATTGGGCCCT CTAGATGCAT GCTCGAGCGG CCGCCAGTGT 60  
 GATGGATATC TGCAGAAATC GCCCTTGTC CGCTGTGCTC CTGATCCAGC GAGGCGCCCA 120  
 TTGCTGCTCC CAATTGGGCT AAAGGCTTGC CATTGTTCCC ACACGGCTAA GTGCCCGGGT 180  
 TCATCCTAAT TGAGCTGAAC ACTAGTCACT GGGTTCCATG GTTCTCTTCC ATGACCCACG 240  
 30 GCTTCTAATA GAGCTCTAAT ACTCACCACA TGGCCCAAGA TTCCATTCTT TGAATCCGT 300  
 GAGGCCAAGA ACCCCAGGTC AGAGAACACG AGGCTTGCCA CGTCTTGGA AGTGGCCCGC 360  
 CGCCATCTTG GGAGCTCTGG GAGCAAGGAC CCCCAGTAA CATTTTGGCA ACCACAAAGG 420  
 GACCTCCAAA GCGATGGGAA ACATTCCTCC CAAGGCAAAA ACGCCCTTAA GATGTATTCT 480  
 GGAGAATTGG GACCAATGTG AACTCAGAC GCTAAGAAAG AAACGATTTA TATTCTTCTG 540  
 35 CAGTACCGCC TGGCCACAAT ATCCTCTTCA AGGGAGAGAA ACCTGGCTTC CTGAGGGAAG 600  
 TATAAATTAT AACATCATCT TACAGCTAGA CCTCTTCTGT AGAAAGGAGG GCAAATGGAG 660

TGAAGTGCCA TATGTGCAAA CTTTCTTTTC ATTAAGAGAC AACTCACAAT TATGTAAAAA 720  
 GTGTGGTTTA TGCCCTACAG GAAGCCCTCA GAGTCCACCT CCCTACCCCA GCGTCCCCC 780  
 CCCGACTCCT TCCTCAACTA ATAAGGACCC CCCTTTAACC CAAACGGTCC AAAAGGAGAT 840  
 AGACAAAGGG GTAAACAATG AACCAAAGAG TGCCAATATT CCCCATTAT GCCCCTCCA 900  
 5 AGCAGTGAGA GGAGGAGAAT TCGGCCCAGC CAGAGTGCCT GTACCTTTT CTCTCTCAGA 960  
 CTTAAAGCAA ATTAAATAG ACCTAGGTAA ATTCTCAGAT AACCTGACG GCTATATTGA 1020  
 TGTTTTACAA GGGTTAGGAC AATCCTTTGA TCTGACATGG AGAGATATAA TGTTACTACT 1080  
 AAATCAGACA CTAACCCCAA ATGAGAGAAG TGCCGCTGTA ACTGCAGCCC GAGAGTTTGG 1140  
 CGATCTTTGG TATCTCAGTC AGGTCAACAA TAGGATGACA ACAGAGGAAA GAACAACTCC 1200  
 10 CACAGGCCAG CAGGCAGTTC CCAGTGTAGA CCCTCATTGG GACACAGAAT CAGAACATGG 1260  
 AGATTGGTGC CACAAACATT TGCTAACTTG CGTGCTAGAA GGACTGAGGA AAACCTAGGAA 1320  
 GAAGCCTATG AATTACTCAA TGATGTCCAC TATAACACAG GGAAAGGAAG AAAATCCTAC 1380  
 TGCTTTTCTG GACAGACTAA GGGAGGCATT GAGGAAGCAT ACCTCCCTGT CACCTGACTC 1440  
 TATTGAAGGC CAACTAATCT TAAAGGATAA GTTATCACT CAGTCAGCTG CAGACATTAG 1500  
 15 AAAAACTTC AAAAGTCTGC CTTAGGCCCG GAGCAGAACT TAGAAACCCT ATTAACTTG 1560  
 GCATCCTCAG TTTTATATA TAGAGATCAG GAGGAGCAGG CGAAACGGGA CAAACGGGAT 1620  
 AAAAAAAAAA GGGGGGTCC ACTACTTTAG TCATGCCCT CAGGCAAGCA GACTTTGGAG 1680  
 GCTCTGGAAA AGGGAAAAGC TGGGCAAATC AAATGCCTAA TAGGGCTGGC TTCCAGTGCG 1740  
 GTCTACAAGG AACTTTTAAA AAAGATTATC CAAGTAGAAA TAAGCGCCCC CCTGTGCCAT 1800  
 20 GCCCCTTAGC TCAAGGGAAT CACTGGAAGG CCCACTGCCC CAGGGGATGA AGATACTCTG 1860  
 AGTCAGAAGC CATTACCAG ATGATCCAGC AGCAGGACTG AGGGTGCCCG GGGCGAGCGC 1920  
 CAGCCCATGC CATCACCTN ACAGAGCCCC GGGTATGCTT GACCATTGAG AGCCAGGAGG 1980  
 TTAAGTGTCT CCTGGACACT GGCGCAGCC 2009

25 (2) INFORMATIONS POUR LA SEQ ID NO:3:

(i) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 1056 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

30 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN.

(xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:3:

CTAGAACGT ATTCTGGAGA ATTGGGACCA ATGTGACACT CAGACGCTAA GAAAGAAACG  
 ATTTATATTC TTCTGCAGTA CCGCTGGCC ACAATATCCT CTTCAAGGGA GAGAAACCTG  
 35 GCTTCCTGAG GGAAGTATAA ATTATAACAT CATCTTACAG CTAGACCTCT TCTGTAGAAA  
 GGAGGGCAAA TGGAGTGAAG TGCCATATGT GCAAACCTTC TTTTCATTAA GAGACAACCT

ACAATTATGT AAAAAGTGTG GTTTATGCCC TACAGGAAGC CCTCAGAGTC CACCTCCCTA  
 CCCCAGCGTC CCCTCCCGA CTCCTTCCTC AACTAATAAG GACCCCTT TAACCCAAAC  
 GGTCCAAAAG GAGATAGACA AAGGGGTAAA CAATGAACCA AAGAGTGCCA ATATTCCCCG  
 ATTATGCCCC CTCCAAGCAG TGAGAGGAGG AGAATTCGGC CCAGCCAGAG TGCCTGTACC  
 5 TTTTCTCTC TCAGACTTAA AGCAAATTAA AATAGACCTA GGTAAATTCT CAGATAACCC  
 TGACGGCTAT ATTGATGTTT TACAAGGGTT AGGACAATCC TTTGATCTGA CATGGAGAGA  
 TATAATGTTA CTAATAATC AGACACTAAC CCCAATGAG AGAAGTGCCG CTGTAAGTGC  
 AGCCCGAGAG TTTGGCGATC TTTGGTATCT CAGTCAGGCC AACAAATAGGA TGACAACAGA  
 GGAAAGAACA ACTCCACAG GCCAGCAGGC AGTTCCTAGT GTAGACCCTC ATTGGGACAC  
 10 AGAATCAGAA CATGGAGATT GGTGCCACAA ACATTGCTA ACTTGCCTGC TAGAAGGACT  
 GAGGAAACT AGGAAGAAGC CTATGAATTA CTCAATGATG TCCACTATAA CACAGGGAAA  
 GGAAGAAAAT CTTACTGCTT TTCTGGACAG ACTAAGGGAG GCATTGAGGA AGCATACCTC  
 CCTGTCACCT GACTCTATTG AAGGCCAACT AATCTTAAAG GATAAGTTTA TCACTCAGTC  
 AGCTGCAGAC ATTAGAAAAA ACTTCAAAAG TCTGCCT

15

## (2) INFORMATIONS POUR LA SEQ ID NO:4:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- 20 (A) LONGUEUR: 25 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN

## 25 (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:4:

CGGACATCCA AAGTGATGGG AAACG

25

## (2) INFORMATIONS POUR LA SEQ ID NO:5:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- 30 (A) LONGUEUR: 26 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN

## 35 (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:5:

GGACAGGAAA GTAAGACTGA GAAGGC

26

## (2) INFORMATIONS POUR LA SEQ ID NO:6:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- (A) LONGUEUR: paires de bases
- 5 (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN

## (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:6:

10 CCT AGA ACG TAT TCT GGA GAA TTG GG

## (2) INFORMATIONS POUR LA SEQ ID NO:7:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- (A) LONGUEUR: paires de bases
- 15 (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN

## (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:7:

20 TGG CTC TCA ATG GTC AAA CAT ACC CG

## (2) INFORMATIONS POUR LA SEQ ID NO:8:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- (A) LONGUEUR: 21 paires de bases
- 25 (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN

## (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:8:

30 ATG GGA AAC GTT CCC CCC GAG

## (2) INFORMATIONS POUR LA SEQ ID NO:9:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- (A) LONGUEUR: 21 paires de bases
- 35 (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:9:

GGC CTA AGG CAG ACT TTT GAA

5

(2) INFORMATIONS POUR LA SEQ ID NO: 10:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: acides aminés

(B) TYPE: acide aminé

10

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: peptide

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 10:

MGSSHHHHHSSGLVP RGSHMASMTGGQQMGR

15 IMGNIIPKAKT PLRCILENWD QCQTQTLRKK RFIFFCSTAW POYPLQGRET

WLPEGSINYN IILQLDLPCR KEGKWSEVPY VQTFFSLRDN SQLCKKCGLC

PTGSPQSPFP YPSVPPPTPS STNKDPPLTQ TVQKEIDKGV NNEPKSANIP

RLCPLQAVRG GEFQPARVPV PFSLSDLKQI KIDLKGFSDN PDGYIDVLQG

LGQSFDLTWR DIMLLNQTL TPNERSAAVT AAREFGDLWY LSQVNNRMTT

20 EERTTPTGQQ AVPSVDPHWD TESEHGDWCH KHLITCVLEG LRKTRKKPMN

YSMMSTITQG KEENPTAFLD RLREALRKHT SLSPDSIEGQ LILKDKFITQ

SAADIRKNFK SLPKLAAALEHHHHHH

(2) INFORMATIONS POUR LA SEQ ID NO: 11:

25

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: acides aminés

(B) TYPE: acide aminé

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

30

(ii) TYPE DE MOLECULE: peptide

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 11:

Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Ile Met Gly Asn

Ile Pro Pro Lys Ala Lys Thr Pro Leu Arg Cys Ile Leu Glu Arg

Ile Leu Glu Asn Trp Asp Gln Cys Asp Thr Gln Thr Leu Arg Lys Lys

35 Arg Phe Ile Phe Phe Cys Ser Thr Ala Trp Pro Gln Tyr Pro Leu Gln

Gly Arg Glu Thr Trp Leu Pro Glu Gly Ser Ile Asn Tyr Asn Ile Ile

Leu Gln Leu Asp Leu Phe Cys Arg Lys Glu Gly Lys Trp Ser Glu Val

Pro Tyr Val Gln Thr Phe Phe Ser Leu Arg Asp Asn Ser Gln Leu Cys

Lys Lys Cys Gly Leu Cys Pro Thr Gly Ser Pro Gln Ser Pro Pro Pro  
 Tyr Pro Ser Val Pro Ser Pro Thr Pro Ser Ser Thr Asn Lys Asp Pro  
 Pro Leu Thr Gln Thr Val Gln Lys Glu Ile Asp Lys Gly Val Asn Asn  
 Glu Pro Lys Ser Ala Asn Ile Pro Arg Leu Cys Pro Leu Gln Ala Val  
 5 Arg Gly Gly Glu Phe Gly Pro Ala Arg Val Pro Val Pro Phe Ser Leu  
 Ser Asp Leu Lys Gln Ile Lys Ile Asp Leu Gly Lys Phe Ser Asp Asn  
 Pro Asp Gly Tyr Ile Asp Val Leu Gln Gly Leu Gly Gln Ser Phe Asp  
 Leu Thr Trp Arg Asp Ile Met Leu Leu Leu Asn Gln Thr Leu Thr Pro  
 Asn Glu Arg Ser Ala Ala Val Thr Ala Ala Arg Glu Phe Gly Asp Leu  
 10 Trp Tyr Leu Ser Gln Ala Asn Asn Arg Met Thr Thr Glu Glu Arg Thr  
 Thr Pro Thr Gly Gln Gln Ala Val Pro Ser Val Asp Pro His Trp Asp  
 Thr Glu Ser Glu His Gly Asp Trp Cys His Lys His Leu Leu Thr Cys  
 Val Leu Glu Gly Leu Arg Lys Thr Arg Lys Lys Pro Met Asn Tyr Ser  
 Met Met Ser Thr Ile Thr Gln Gly Lys Glu Glu Asn Leu Thr Ala Phe  
 15 Leu Asp Arg Leu Arg Glu Ala Leu Arg Lys His Thr Ser Leu Ser Pro  
 Asp Ser Ile Glu Gly Gln Leu Ile Leu Lys Asp Lys Phe Ile Thr Gln  
 Ser Ala Ala Asp Ile Arg Lys Asn Phe Lys Ser Leu Pro Lys Leu Ala  
 Ala Ala Leu Glu His His His His His His

20 (2) INFORMATIONS POUR LA SEQ ID NO:12:

(i) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 26 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

25 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:12:

CTAGAACGTA TTCTGGAGAA TTGGGA

26

30 (2) INFORMATIONS POUR LA SEQ ID NO:13:

(i) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 20 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

35 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:13:

CCTAAGGCAG ACTTTTGAAG

20

(2) INFORMATIONS POUR LA SEQ ID NO:14:

5 (i) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 54 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

10 (ii) TYPE DE MOLECULE: ADN.

(xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:14:

TTTGTAATA CGACTCACTA TAGGGCAGCC ACCATGGGAA ACGTTCCCCC CGAG

54

(2) INFORMATIONS POUR LA SEQ ID NO:15:

15 (i) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 44 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

20 (ii) TYPE DE MOLECULE: ADN.

(xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:15:

TTTTTTTTTTT TTTTTTTTTC AGGCTGCGCC AGTGTCCAGG AGAC

(2) INFORMATIONS POUR LA SEQ ID NO:16:

25 (i) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 20 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

30 (ii) TYPE DE MOLECULE: ADN.

(xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:16:

TGTCCGCTGT GCTCCTGATC

20

(2) INFORMATIONS POUR LA SEQ ID NO:17:

35 (i) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 44 paires de bases

- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

5 (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:17:

TTTTTTTTTT TTTTTTTTTC AGGCTGCGCC AGTGTCCAGG AGAC

44

(2) INFORMATIONS POUR LA SEQ ID NO:18:

(i) CARACTERISTIQUES DE LA SÉQUENCE:

- 10 (A) LONGUEUR: 678 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

15 (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:18:

TCAGGGATAGCCCCATCTATTTGGCCAGGCATTAGCCCAAGACTTGAGCCAGTTCTCATACCTGGATAT  
TCTTGTCTCTTTGGTATGCGGATGATTTACTTTTAGCGCCCCGTTTCCAGAAACCTTGCGCATCAAGCCACC  
CAAGTGTCTTAAATTTCTCGCCACCTGTGGCTACAAGGTTTCCAAACCAAGGCTCAGCTCTGCTCAC  
AGCAGAAGGCTATTTACCCTAAATACTTAGGGCTGAAATTATCCAAAGGCACCAGGGCCCTCAGTGAGGA  
20 ATGTATCCAGCCTATACTGGCTTATCCTTATCCCAAAACCTTAAACAACTAAGAAGGTTCTTGGCATA  
ATAGGCATAACAGGCATAACAGGTTTCTGCTGAATATGGATTCCCAAGTAOGGCAAAATAGCCAGACCAT  
TATATACACTAATTAAGGAACTCAGAAAGCCAATACCCATTTAGTAAGATGGACACCTGAAGCAGAGGC  
AGCTTTCCAGGCCGTAAAGAACCCCTAACCCCAAGCCCCAGTGTTAAGCTTGCCAGCGGGCAAGACTTT  
TCTTTCTGTGTACAGAAAAATAGGAATAGCTNTAGGAGTCCTTACACAGGTCGAGGGACCAGCTTGC  
25 AACCCATGGCATACTGAGTAAGGAAATTGATGTAGTGGCAAAGGGTT

(2) INFORMATIONS POUR LA SEQ ID NO:19:

(i) CARACTERISTIQUES DE LA SÉQUENCE:

- 30 (A) LONGUEUR: 591 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:19:

35 CCATGGCCATCTACACTGAACAAGATTTATACAATCATGTCGTACCTAAGCCCCACAACAAAGAGTACC  
CATTCTTCCTTTTGTATCAGAGCAGGAGTGCTAGGCAGACTAGGTACTGGCATTGGCAGTATCACAACC  
TCTACTCAGTTCTACTACAACTATCTCAAGAAATAAATGGTGACATGGAACAGGTCACTGACTCCCTGG  
TCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAATCGAAGAGCTTTAGACTTGCT  
AACCGCCAAAGAGGGGGAACCTGTTTATTTTAGGAGAAGAAGCTGTTATTATGTTAATCAATCCAGA  
40 ATTGTCACTCAGAAAGTTAAAGAAATTCGAGATCGAATACAATGTAGAGCAGAGGAGCTTCAAACACCG  
AACGCTGGGGCCTCCTCAGCCAATGGATGCCCTGGGTCTCCCTTCTTAGGACCTCTAGCAGCTCTAAT  
ATTGTTACTCTCTTGGACCCTGTATCTTTAACCTCCTTGTAAAGTTTGTCTCTTCCAGAATTGAAGCT  
GTAAGCTACAGATGGTCTTACAAATCTAGA

## (2) INFORMATIONS POUR LA SEQ ID NO:20:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 1321 paires de bases

5 (B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN

## (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:20:

10 CAACAATCGGGATATAAACCAGGCATTGAGCTGGCAACAGCAGCCCCCTTTGGGTCCCTTCCCTTTG  
TATGGGAGCTGTTTTTCATGCTATTTCACTCTATTAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCT  
TACGGCTCGAGCTGAGCTTTTGCTCACCGTCCACCCTGCTGTTTGCCACCACCGCAGACCTGCCGCTGA  
CTCCCATCCCTCTGGATCCTGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAAGCGCCATTGCCGCTCC  
CAATTGGGCTAAAGGCTTGCCATTGTTCTGACGCTAAGTGCCCTGGGTTTGTCTAATTGAGCTGAAC  
15 ACTAGTCACTGGGTTCATGGTCTCTTCTGTGACCCACGGCTTCTAATAGAACTATAACACTTACCACA  
TGGCCCAAGATTCCATTCTTGGAAATCCGTGAGGCCAAGAACTCCAGGTGAGAGAAACGAAGCTTGCCA  
CCATCTTGGAAGCGGCCTGTACCATCTTGGAAGTGGTTCAACCATCTTGGGAGCTCTGTGAGCAAGG  
ACCCCCGGTAACATTTTGGCAACCAAGACGACATCCAAAGTGATGGGAAACGTTCCCGCAAGACAA  
AAACGCCCTAAGACGTATTCTGGAATTTGGGAACAATTTGACCTCAGACACTAAGAAAGAAACGACT  
20 TATATTCTTCTGAGTGCCGCTGGCACTCCTGAGGGAAGTATAAATTATAACACCATCTTACAGCTAGA  
CCTCTTTTGTAGAAAAGGCAATGGAGTGAAGTGCCATAAGTACAACTTTCTTTTCTAATAGAGACAAC  
TCACAATTATGTAAAAGTGTGATTTATGCCCTACAGGAAGCCTTCAAGTCTACCTCCCTATCCAGCA  
TCCCGGACTCCTTCCCACTTAATAAGGACCCCCCTTCAACCCAAATGGTCCAAAAGGAGATAGACAAA  
GGGTAAACAGTGAACCAAGAGTGCCAAATATCCCAATTATGACCCCTCAAGCAGTGGGAGGAAGAGA  
25 ATTCGGCCAGCCAGAGTGCATGTCCCTTTTCTCTCCAGACTTAAAGCAAATAAAACAGACTTAGGT  
AAATTCTCAGATAACCTGATGGCTATATTGGTGTTTTACAAGGGTTAGGACAATTCTTTGATCTGACAT  
GGAGAGATATATATGCTACTGCTAAATCAGACACTAACCCCAATGAGAGAAGTGCCACCATAACTGCAG  
CCTGAGAGTTTGGCGATCTCTGGTATCTCAGTCAGGTCAATGATAGGATGACAACAGAGGA

## 30 (2) INFORMATIONS POUR LA SEQ ID NO:21:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 2938 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

35 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN

## (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:21:

CAACGACGGACATCCAAAGTGATGGGAAACGTTCCCGCAAGACAAAAACGCCCTAAGACGTATTCTGG  
AGAATTGGGACCAATTGACCCTCAGACACTAAGAAAGAAACGACTTATATTCTTCTGCACTGCCGCTG  
40 GCACTCCTGAGGGAAGTATAAATTATAACACCATCTTACAGCTAGACTTCTTTGTAGAAAAGGCAAATG  
GAGTGAAGTGCCATAAGTACAACTTTCTTTTCTAATAGAGACAACTCACAATTATGTAAGTGTGAT  
TTATGCCCTACAGGAAGCCTTCAAGTCTACCTCCCTATCCAGCATCCCGACTCCTTCCCCAACTAAT  
AAGGACCCCTTCAACCAATGGTCCAAAAGGAGATAGACAAAAGGTAACAGTGAACCAAGAGTG  
CCAATATTCCCAATTATGACCCCTCCCAAGCAGTGGGAGGAAGAGATTCCGCCCAGCCAGAGTGCATGT  
45 GCTTTTCTTCTCCAGACTTAAAGCAAATAAAACAGACTTAGGTAAATTCTCAGATAATCCTGATGGC  
TATATTGATGTTTACAAGGGTTAGGACAATTCTTTGATCTGACATGGAGAGATATAATGTCACTGCTAA  
ATCAGACACTAACCCCAATGAGAGAAGTGCCACCATAACTGCAGCTGAGAGTTTGGCGATCTCTGGTA

TCTCAGTCAGGTCAATGATAGGATGACAAACAGAGGAAAGAGATGATCCCCACAGCCAGCAAGCAGTTCCC  
 AGTCTASACCCTCATTGGGGACACAGAAATCAGTAACATGGGAGATTGGTGCTGCAGACATTGCTAACT  
 TGTGTGCTACAAGGACTAAGGAAAACACGAAGAAAATCTACGAATTACTCAATGATGTCCACCATAACA  
 CAGGGGAAGGGAAGAAAATCCTACTGCCTTTCTGGAGAGACTAAGGGAGGCATTGAGGAAGCGTGCCTCT  
 5 CTGTCACTGACTCTTCTGAAGGCCAACTAATCTTAAAGCGTAAGTTTATCACTCAGTCAGCTGCAGACA  
 TTAGAAAAAACTTCAAAGTCTGCCGTAGGCCCGGAGCAAACTTAGAAACCTATTGAACCTTGGCAACY  
 TCGGTTTTTTTATAATAGAGATCAGGAGGAGCAGCGCGGAACAGGACAAACGGGATTAAAAAAAAGGCCACC  
 GCTTTAGTCATGACCCTCAGGCAAGTGGACTTTGGAGGCTCTGGAAAAGGGAAGCTGGGCAAAATTGAA  
 TGCCTAATAGGGCTTCTTCCAGTGGCTCTACAAGGACACTTTAAAAAGATTGTCCAAGTAGAAGTAA  
 10 GCCGCCCTTCGTCCATGCCCTTATTTCAAGGAATCACTGGAAGGCCACTGCCCCAGGGGACAAAGG  
 TCTTTTGTAGTCAGAAGCCACTAACAGATGATCCAGCAGCAGGACTGAGGGTGCCTGGGGCAAGCGCCAT  
 CCCATGCCATCACCCTCAGAGGCCCTGGGTATGCTTGACCATTGAGGGCCAGGAAGGTTGCTCCTGGA  
 CACTGGTGGCGTCTTCTTAGTCTTACTCTTCTGTCCCGGACAACTGTCTCCAGATCTGTCTACTATCTGA  
 GGGGCTCCTAAGACGGGCAGTCACTAGATACTTCTCCAGCCACTAAGTTATGACTGGGGAGCTTTATTC  
 15 TTTTCACATGCTTTTCTAATTATGCTTGAAGCCCCACTACCTTGTAGGGAGAGACATTCTAGCAAAAG  
 CAGGGGCCATTATACACCTGAACATAGGAGAAGGAACACCCGTTGTTGTCCCCTGCTTGAGGAAGGAAT  
 TAATCCTGAAGTCTGGGCAACAGAAGGACAATATGGACGAGCAAGAATGCCCGTCTGTTCAAGTTAAA  
 CTAAGGATTCCACTTCCTTTCCCTACCAAGGCAGTACCCCTCAGACCCAAGGCCCAACAGGATTCC  
 AAAAGATTGTTTAAAGACTTAAAGCCCAAGGCTTAGTAAACCATGCATAACTCCCTGCAGTAATTCCGT  
 20 AGTGGATTGAGGAGGCACAGAAACCCAGTGGACAGTGGAGGCTTAGTGCAAGATCTCAGGATTATCAATG  
 GAGGCCGTTGCTTTTATACCCAGCTGTACCTAGCCCTTACTGTGCTTTCCCAATACCAGAGGAAG  
 CAGAGTGGTTTACACTCCTGGACCTTAAGGATGCCTTCTTCTGCATCCCTGTACATCCTGACTCTCAATT  
 CTTGTTTGCCTTTGAAGATACTTCAAAACCAACATCTCAACTCACCTGGACTGTTTACCCCAAGGGTTC  
 AGGGATAGCCCCATCTATTTGGCCAGGCATTAGCCCAAGACTTGAGCCAATCCTCATACCTGGACACTT  
 25 GTCCTTGGTAGGTGGATGATTTACTTTGGCCGCCCATTCAGAAACCTTGTGCCATCAAGCCACCCAAAG  
 CGCTCTTCAATTTCTCGCTACCTGTGGCTACATGGTTTCCAAACCAAGGCTCAACTCTGCTCAGAGCA  
 GGTACTTAGGGCTAAAATTATCCAAAGGCACCAAGGCCCTCAGTGAGGAACACATCCAGCCTATACTGG  
 CTTATCCTCATCCCAAAACCTAAAGCAACTAAGGGGATTCTTGGCGTAATAGGTTTCTGCCGAAATG  
 GATTCCCAGGTTTGGCGAATAGCCAGGTCATTAAATACACTAATTAAGGAAACTCAGAAAGCCATACC  
 30 CATTAGTAGATGGAACACTGAAGTAGAAGTGGCTTTCCAGGCCCTAACCAAGCCCCAGTGTTAAGTT  
 TGCCAAACAGGGCAAGACTTTCTTCATATGTACAGAAAAACAGGAATAGCTCTAGGAGTCCCTACACA  
 GATCCGAGGGATGAGCTTGCAACCTGTGGCGTACCTGACTAAGGAAATTGATGTAGTGGCAAGGGTT

(2) INFORMATIONS POUR LA SEQ ID NO:22:

35 (i) CARACTERISTIQUES DE LA SÉQUENCE:

- (A) LONGUEUR: 1422 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

40 (ii) TYPE DE MOLECULE: ADN.

(xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:22:

TCAGGGATAGCCCCATCTATTTGGCCAGGCATTAGCCCAAGACTTGAGTCAGTTATCATACCTGGACAC  
 TCTTGTCCTTCAGTATGTGGATGATTTACTTTAGCTGCCTGTTCAGAAACCTTGTGCCATCAAGCCACC  
 CAAGCACTCTTAAATTTCTCGCCACCTGTGGCTACAAGGTTTCCAAAGAGAAGCTCAGCTCTGCTCACA  
 45 GCAGGTTAAATACTTAGGACTAAGATTATCCAAAGGCACCAAGGCCCTCAGTGAGGAATGTATCCAGCCT  
 ATACTGGCTTATCCTCATCTCAAAACCTTAAGCAACTAAGAGAGTTCTTGGCATAACAGGCTTCTGCC  
 GAATATGGATTCCCAAGGTATGGCAAAATAGCCAGGCCATTATATACAGTAATTAAGGAAACTCAGAAAG  
 CCAATACCCATTTAATAAGATGGATACCTGAAGCCAAAGTGGCTTTCCAGGCCCTAAAGAAGGCCCTTAA  
 ACCCAAGTCCCAAGTGAAGCTTGCCAACGGGCAAGACTTTCTTTATACATCAGAAAAAAACAGAA  
 50 ACAGCTCTGGGAGTCCCTTACACAGTCCCAAGGGACGAGCTTGCAACCCATGGCATACTGAGTAAGGAAA  
 CTGATGTAGTGGCAAGGGTTGGCTTCATTGTTTATGGGTAGTGGTGGCAGTAGCAGTTGTAGTATCTGA  
 AGCAGTTAAATAATACAGGGGAGAGATCTTACTGTGTGGACATCTCATGAGGTGAACAGCATACTCACT

GCTAAAGGAGACTTGTGGCTGTGACACACCGTTTACTTAAATATCAGGCTCTATTACTTGAAAGGCCAG  
 TGCTGCAACTGTGCACTTGTGCACTCTTAACCCAGTCNCATTTCTTCCAGACATGAAGATAGAATATA  
 ACTGTCAACAAATAATTTCTCAAACCTATGCCACTCGAGGGGACCTTCTAGAAGTTCCCTTGACTGATCC  
 5 TGACCTTCAACTTGTATACTGATGGAAGTTCTTTGTAGAAAAAGGACTTCAAAAGCGGGGTATGCAGTG  
 GTCAGTGATAATGGAATATTTGAAAGTATCCCCCTCACTCCAGGAACTAGTGCTTAGCTGGCAGAACTAAT  
 AGCCTTCATTGGGGCACTAGAATTAGGAGAAGGAAAAAGGGTAAATATATATACAGACTCTGAGTATGCT  
 CACCTAGTCNTCCATGCCCATGAGGCAATATGCAGAGAAAGGGAATTCCTAACTTCCGAGGGAACACCTA  
 TCACACATCAGGAAGCCATTAGGAGATTATTACTGGCAGTACAGAAACCTAAAGAGGTGGAAGTCTTACA  
 10 CTGCTGGGGTCATCAGAAAGGAAAGAAAAGGGAATAGAAGGGAATTGCCAAGCAGATATTGAAGCAAAA  
 AGAGCTGCAAGGCAGGACCCTC

(2) INFORMATIONS POUR LA SEQ ID NO:23:

(i) CARACTERISTIQUES DE LA SÉQUENCE:

- (A) LONGUEUR: 2006 paires de bases  
 15 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:23:

20 ATGCAGTGGTCAGTGATAATGGAATACTTGAAGTAATCCCCTCACTCCAGGAACTAGTGCTCAGCTAGC  
 AGAACTAATAGCCCTCACTTGGGCACTAGAATTAGGAGAAGAAAAAGGGCAAATATATATACAGACTCT  
 AAATATGCTTACCTAGTCTCCATGCCCCATGCAGCAATATGGAAGAAAGGGAATTCCTAACTTCTGAGA  
 GAACACCTATCAACATCAGGAAGCCATTAGGAATTATTATTGGCTGTACAGAAACCTAAAGAGGTGGC  
 25 AGTCTTACACTGCCGGGTTCATCANAAGGAAAGGAAAGGGAATACTTTTGCTGCAACTATCCAATG  
 GAAATTACTTAAACCCCTTCATCAAAACCTTTCCTTAGGCATCGATAGCACCCATCAAATGGCCAAATCA  
 TTATTTACTGGACCAGGCCTTTTCAAACCTATCAAGCAAATATTCAGGGCTGTGAATTGTGCCAAAAAA  
 ATAATCCCCTGCCTCATGCCAAGCTCCTTCAGGAAAACAAAAACAGGCCATTACCCTGAAAAAACTG  
 GCAACTGATTTTACCCACAAGCCCAAACCTCAGGGATTTCAGTATCTACTAGTCTGGGTAAATACTTTCA  
 CGGGTTGGGCAAGGCCTTCCCCTGTAGGACAGAAAAGGCCAAGAGGTAATAAAGGCACTAGTTCATGA  
 30 AATAATCCCAGATTCCGACTTCCCGAGGCTTACAGAGTGACAATAGCCCTGCTTTCAGGCCACAGTA  
 ACCCAGGGAGTATCCAGGCGTTAGGTATACGATACCTTACACTGCGCCTGAAGGCCACAGTCCTCAG  
 GGAAGGTCGAGAAAATGAATGAATACTCAAAGGACATCTAAAAAGCAAACCCAGGAAACCCACCTCAC  
 ATGGCCTGCTCTGTTGCTTATAGCCTTAAAAAGAATCTGCAACTTCCCCTAAAAAGCAGGACTTAGCCCA  
 TACGAAATGCTGTATGGAAGGCCCTTCATAACCAATGACCTTGTGCTTGACCCAAGACAGCCAACTTAGT  
 35 TGCAGACATCACCTCCTTAGCCAAATATCAACAAGTTCTTAAACATTACAAGGAACCTATCCCCTGAGAA  
 GAGGGAAGAACTATTCCACCCTTGTGACATGGTATTAGTCAAGTCCCTTCTCTTAATTCCCATCCC  
 TAGATACATCCTGGGAAGGACCTACCCAGTCATTTTATTTACCCCAACTGCGGTTAAAGTGGCTGGAGT  
 GGTCTTGATACATCACACTTGAGTCAAATCCTGGTACTGCCAAAGGAACCTGAAAAATCCAGGAGACAA  
 CGCTAGCTATTCTGTGAACCTCTAGAGGATTGCGCCTGCTCTTCAACAACAACCCAGGAGGAAAGTAA  
 40 CTAAATCATAAATCCCCATGGCCCTCCCTTATCATATTTTCTCTTACTGTTCTTTACCCCTTTTC  
 ACTCTCACTGCAACCCCTCCATGCCGCTGTATGACCAGTAGCTCCCCTTACCAAGAGTTTCTATGGAGAA  
 TGCAGCGTCCCAGAAATATTGATGCCCATCGTATAGGAGTCTTTCTAAGGGAACCCCACTTCACTGCG  
 CCACACCCATATGCCCGCAACTGCTATCACTCTGCCACTCTTGCATGCATGCAATACTCATTATTGG  
 ACAGGAAAAATGATTAATCCTAGTTGCTGAGGAGTGGAGTCACTGTCTGTTGGACTTACTTCACCC  
 45 AAAGTGGTATGCTGATGGGGGTGGAGTTCAAGATCAGGCAAGAGAAAAACATGTAAAGAAAGTAATCTC  
 CCAACTCACCCGGGTACATGGCACCTCTAGCCCTACAAAGGACTAGATCTCTCAAACTACATGAACCCC  
 TCCGTACCCATACTGCGCTGGAAGCCTATTAAATACCACCCTCACTGGGCTCCATGAGGTCTCGGCCCA  
 AAACCTACTAACTGTTGGATATGCCTCCCCTGAACTTCAAGCCA

50 (2) INFORMATIONS POUR LA SEQ ID NO:24:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- (A) LONGUEUR: 1948 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

5

## (ii) TYPE DE MOLECULE: ADN

## (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:24:

ACTGCACCTTTCTGGTCCATGTTTCTTACGGCTCGAGCTGAGCTTTTGCTCACCGTCCACCACTGCTGTT  
 TGCCACCACCGCANACCTGCCGCTGACTCCCATCCCTCTGGATCCTGCAGGGTGTCGCTGTGCTCCTGA  
 10 TCCAGCGAGGCGCCATTGCCGCTCCCAATTGGGCTAAAGGCTTGCCATTGTNCTGCACGGCTAAGTGC  
 CTGGGTTTGTCTAATTGAGCTGAACACTANTCACTGGGTTCCATGGTTCTCTTCTGTGACCCACGGCTT  
 CTAATAGAACTATAACACTTACCACATGGCCCAAGATTCCATTCCCTTGGAAATCCGTGAGGGCAAGAACTC  
 CAGGTGAGAGAATACGAGGCTTGCCACCATCTTGGAGCGGCTGCTACCATCTTGGAAAGTGGTTACCA  
 CCATCTTGGGAGCTCTGTGAGCAAGGACCCCGGTAACATTTGGCAACCACGAACGGACATCCAAAGT  
 15 GATACATCCTGGGAAGGACCTACCCAGTCATTTTATCTACCCCACTGCGGTAAAGTGGCTGGAGTGG  
 AGTCTTGGATACATCACTTGAGTCAAATCCTGGATACTGCCAAAGGAACTGAAAATCCAGGAGACAA  
 CGCTAGCTATTCTGTGAACCTCTAGAGGATTTGCGCCTGCTCTTCAAACAACACCAGGAGGAAAGTAA  
 CTAAAATCATAAATCCCCATGGCCCTCCCTTATCATTTTTCTCTTACTGTGTTTCAACCTCTTTCA  
 CTCTCACTGCACCCCTCCATGCCGCTGTATGACCAGTAGCTCCCTTACCAAGAGTTTCTATGGAGAAT  
 20 GCAGCGTCCCGGAAATATTGATGCCCCATCGTATAGGAGTCTTTGTAAGGGAACCCACCTTCACTGCC  
 CACACCCATATGCCCGCACTGCTATCACTCTGCCACTCTTGCATGCGATGCAAATACTCATTATTGGA  
 CAGGAAAAATGATTAATCCTAGTTGTCTGGAGGACTTGGAGTCACTGTCTGTTGGACTTACTTCACCCA  
 AACTGGTATGTCTGATGGGGTGGAGTTCAAGATCAGGCAAGAGAAAAACATGTAAGAAGTAATCTCC  
 CAACTCACCCGGGTACATGCCACCTCTAGCCCTACAAAGGACTAGATCTCTCAAACTACATGAAACCC  
 25 TCCGTACCCATACTCGCCTGGTAAGCCTATTTAATACCACCTCACTGGGCTCCATGAGGTCTCGGCCCA  
 AAACCTTACTAAGTTGGATATGCCTCCCGCTGAACCTCAGGCCATATGTTTCAATCCCTGTACCTGAA  
 CAATGGAAACAACCTCAGCACAGAAATAAACACCACTTCCGTTTTAGTAGGACCTCTTGTCTTCCAACTG  
 AAATAACCCATACCTCAAACCTCACCTGTGTAATAATTTAGCAATACTACATACAAACCACTCCCAATG  
 CATCAGGTGGGTAACCTCTCCACACAAATAGTCTGCCTACCTCAGGAATATTTTTGTCTGTGGTACC  
 30 TCAGCCTATCGTTGTTTGAATGGCTCTTCAGAATCTATGTCTTCTCTCATCTTAGTGCCCCCTATGG  
 CCATCTACACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACCCATTCT  
 TCCTTTTGTATAGGAGCAGGAGTGTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACT  
 CAGTTCTACTACAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTGCGCGACTCCCTGGTCACCT  
 TGCAAGATCAACTTAACCTCCCTAGCAGCAGTAGTCCTTCAAATCGAAGAGCTTTAGACTTGCTAACCGC  
 35 TGAAAGAGGGGGAACCTGTTATTTTAGGGGAAGAATGCTGTTATTATGTTAATCAA

## (2) INFORMATIONS POUR LA SEQ ID NO:25:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- (A) LONGUEUR: 1136 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

40

## (ii) TYPE DE MOLECULE: ADN

## (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:25:

CCATGGCCATCTACACTGAACAAGATTTATACAGTTATGTCATATCTAAGCCCCGCAACAAAAGAGTACC  
 45 CATTCTTCTTTTGTATAGGAGCAGGAGTGCTAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACC  
 TCTACTCAGTTCTACTACAACTATCTCAAGAACTAAATGGGGACATGGAACGGGTGCGCGACTCCCTGG

TCACCTTGCAAGATCAACTTAACTCCCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTCGCT  
 AACCGCTGAAAGAGGGGGAACCTGTTTATTTTAGGGGAAGAATGCTGTTATTATGTTAATCAATCCGGA  
 ATCGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAACGTAGAGCAGAAGAGCTTCGAAACACTG  
 GACCTGGGGCCTCCTCAGCCAATGGATGCCCTGGATTCTCCCCTTCTTAGGACCTCTAGCAGCTATAAT  
 5 ATTGCTACTCCTCTTTGGACCTGTATCTTTAACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCT  
 GTAAACTACAAATGGAGCCCAAGATGCAGTCCAAGACTAAGATCTACCGCAGACCCCTGGACCGGCTG  
 CTAGCCACGATCTGATGTTAATGACATCAAAGGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCT  
 ACTACGCCCCAATTGAGCAGGAAGCAGTTAGAGCGGTCTGCGGCCAACCTCCCCAACAGCACTTAGGTTT  
 TCCTGTTGAGATGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGCTGACTAAGAATCCCTAAGCC  
 10 TAGCTGGGAAGGTGACCACATCACCTTTAAACACGGGGCTTGCAACTTAGTTACACCTGACCAATCAG  
 AGAGCTCACTAAATGCTAATTAGGCAAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCATGAGA  
 GCACAGCAGGAGGACAATGATCGGGATATAAACCCAAAGTCTTCGAGCGGCAACGGCAACCCCTTTGG  
 GTCCCCCTCCCTTTGTATGGGAGCTCTGTTTTCATGCTATTTCACTCTATTAAATCTTGACGCTGCGAAA  
 AAAAAAAAAAAAAA

15

## (2) INFORMATIONS POUR LA SEQ ID NO:26:

## (1) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 2782 paires de bases

(B) TYPE: nucléotide

20 (C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(1i) TYPE DE MOLECULE: ADN

## (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:26:

ATGGGAGCTGTTTTATGCTATTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTT  
 25 ACGGCTCGAGCTGAGCTTTTGCTCACCCTCCACCCTGCTGTTTGCCACCACCGCAGACCTGCCGCTGAC  
 TCCCATCCCTCTGGATCCTGCAGGGTGTCCGCTGTGCTCCTGATCCAGCGAAGCGCCATTGCCGCTCCC  
 AATTGGGCTAAAGGCTTGCCATTGTTCTGCAACGGCTAAGTGCTGGGTTTGTCTAATTGAGCTGAACA  
 CTAGTCACTGGGTTCCATGGTTCTCTTCTGTGACCCACGGCTTCTAATAGAATATAACACTTACCACAT  
 GGGCCAAGATTCCATTCCCTTGAATCCGTGAGGCCAACGAACTCCAGGTGAGAGAATACGAAGCTTGCCA  
 30 CCATCTTGAAGCGGCTGCTACCATCTTGAAGTGGTTACCACCATCTTGGGAGCTCTGTGAGCAAGG  
 ACCCCCCGGTGACATTTGGCGACCACCAACGGACATCCCAAGTGATACATCCTGGGAAGGACCCCTACCC  
 AGTCATTTTATCTACCCCAACTGCGGTTAAAGTGGCTGGAGTGGAGTCTTGATACATCACACTTGAGTC  
 AAATCCTGGATACTGCCAAAGGAACCTGAAATCCAGGAGACAACGCTAGCTATTCTGTGAACCTCTAG  
 AGGATTTGCGCCTGCTCTTCAACAACAACAGGAGGAAAGTAACTAAATCATAAATCCCCATGGGCT  
 35 CCCTTATCATATTTTCTCTGTAGTGTTCTTACCCTGTTCACTCTCACTGCACCCCTCCATGCCGC  
 TGTATGACCAGTAGCTCCCTCACCAGAGTTTCTATGGAGAATGCAGCGTCCCGGAAATATTGATGCCC  
 CATCGTATAGGAGTCTTCTAAGGGAACCCCACTTCACTGCCACACCCATATGCCCCGAACTGCTA  
 TCACTCTGCCACTCTTTGCATGCATGCAATACTCATTATTGGACAGGAAAAATGATTAATCCTAGTTGT  
 CCTGGAGGACTTGGAGTCACTGTCTGTTGGACTTACTTACCCAACTGGTATGTCTGATGGGGTGGAG  
 40 TTCAAGATCAGGCAAGAGAAAAACATGTAAAGAAAGTAATCTCCCAACTCACCAGGATACATGGCACCTC  
 TAGCCCTACAAAGGACTAGATCTCTCAAACTACATGAAACCTCCGTACCCATACTCGCCTGGTAAGC  
 CTATTTAATACCACCTCACTGGGCTCCATGAGGTCTCGGCCAAAACCTACTAACTGTTGGATATGCC  
 TCCCCCTGAACCTCAGGCCATATGTTCAATCCCTGTACCTGAACAATGGAACAACCTCAGCACAGAAAT  
 AAACACCACTTCCGTTTAGTAGGACCTCTGTTTCCAATGTGGAATAACCCATACCTCAAACCTCACC  
 45 TGTGTAATTTAGCAATACTACATACACAACCAACTCCCAATGCATCAGGTGGGTAACCTCTCCACAC  
 AAATAGTCTGCCTACCTCAGGAATATTTTGTCTGTGGTACCTCAGCCTATCGTTGTTGAATGGCTC  
 TTCAGATCTATGTGCTTCTCTCATTCTTAGTGCCCTATGACCATCTACACTGAACAAGATTTATAC  
 AGTTATGTCTATCTAAGCCCCGCAACAAAGAGTACCCATTCTCTTTTGTATAGGAGCAGGAGTGC  
 TAGGTGCACTAGGTACTGGCATTGGCGGTATCACAACCTCTACTCAGTTCTACTACAACTATCTCAAGA  
 50 ACTAAATGGGACATGGAACGGGTGCGCGACTCCCTGGTCACCTTGCAAGATCACTTAACTCCCTAGCA  
 GCAGTAGTCCTTCGAAATCGAAGAGCTTTAGACTTGCTAACCCTGAGAGAGGGGGAACCTGTTTATTTT  
 TAGGGGAAGAATGCTGTTATTATGTTAATCAATCCGAATCGTCACTGAGAAAGTTGAAGAAATTCAGA

TCGAATACAACGTATAGCAGAGGAGCTTCGAAACACTGGACCCTGGGGCCTCCTCAGCCGATGGATGCCC  
 TGGATTCTCCCCCTTCTTAGGACCTCTAGCAGCTATAATATTGCTACTCCTCTTTGGACCCTGTATCTTTG  
 ACCTCCTTGTTAACTTTGTCTCTTCCAGAATCGAAGCTGTGAAACTACAAATGGAGCCCAAGATGCAGTC  
 CAAGACTAAGATCTACCGCAGACCCCTGGACCGGCTGCTAGCCCACGATCTGATGTTAATGACATCAA  
 5 GGCACCCCTCCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCAATTACAGCAGGAAGCAGTTAGA  
 GCGGTGGTCGGCCAACCTCCCCAACAGCACTTAGGTTTTCTGTTGAGATGGGGGACTGAGAGACAGGAC  
 TAGCTGGATTTCCTAGGCTGACTAAGAATCCTTAAGCCTAGGTGGGAAGGTGACCACATCCACCTTTAAA  
 CACGGGGCTTGCAACTTAGCTCACACCTGACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAAGAC  
 AGGAGGTAAGAAATAGCCAATCATTTATTGCCTGAGAGCACAGCAGGAGGGACAATGATCGGGATATAA  
 10 ACCCAAGTTTTTCGAGCCGGCAACGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTCTGTTTTT  
 ATGCTATTTCACTCTATTAAATCTTGCAACTGCAAAAAAAAAAAAAAAAAAAAAA

## (2) INFORMATIONS POUR LA SEQ ID NO:27:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- 15 (A) LONGUEUR: 666 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN

## 20 (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:27:

TGTCCGCTGTGCTCCTGATCCAGCGAGGCGCCCATTCGCCCTCCCAATTGGGCTAAAGGCTTGCCATTGT  
 TCCTGCACGGCTAAGTGCCTGGGTTTGTCTAATTGAGCTGAACACTANTCACTGGGTCCATGGTTCTC  
 TTCTGTGACCCACGGCTTCTAATATACTATAACACTTACCACATGGCCCAAGATTCCATTCCTTGGAA  
 25 CCGTGAGGCCAAGAACTCCAGGTCAGAGAATACGAGGCTTGCCACCATCTTGAAGCGGCCTGCTACCAT  
 CTTGGAAGTGGTTCAACCACATCTTGGGAGCTCTGTGAGCAAGGACCCCCGGTAACATTTTGGCAACCA  
 CGAACGGACATCCAAAGTGAATCGAAGCTGTAAACTACAAATGGAGCCCAAGATGCAGTCCAAGACTAA  
 GATCTACCGCAGACCCCTGGACCGGCTGCTAGCCCACGATCTGATGTTAATGACATCAAAGGCACCCCT  
 CCTGAGGAAATCTCAGCTGCACAACCTCTACTACGCCCAATTACAGCAGGAAGCAGTTAGAGCGGTCTC  
 GGCCAACCTCCCCAACAGCACTTAGGTTTTCTGTTGAGATGGGGGACTGAGAGACAGGACTAGCTGGAT  
 30 TTCTTAGGCTGACTAAGAATCCCTAAGCCTAGCTGG

## (2) INFORMATIONS POUR LA SEQ ID NO:28:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- 35 (A) LONGUEUR: 3372 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN.

## (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:28:

40 GACTTCCCAA ATACCAGAGG AAGCAGAGTG GTTTACAGTC CTGGACCTTC AGGATGCCTT 60  
 CTTCTGCATC CCTGTACATC CTGACTCTCA ATTCTTGTTT GCCTTTGAAG ATACTTCAA 120  
 CCCAGCATCT CAACTCACCT GGACTATTTT ACCCCAAGGG TTCAGGGATA GTCCCCATCT 180

ATTTGGCCAG GCATTAGCCC AAGACTTGAG CCAATCCTCA TACCTGGACA CTTGTCCTTC 240  
 GGTAGGTGGA TGATTACTT TTGGCCGCCC ATTCAGAAAC CTTGTGCCAT CAAGCCACCC 300  
 5 AAGCGCTCTT CAATTTCCTC GCTACCTGTG GCTACATGGT TTCCAAACCA AAGGCTCAAC 360  
 TCTGCTCACA GCAGGTACT TAGGGCTAAA ATTATCCAAA GGCACCAGGG CCCTCAGTGA 420  
 10 GGAACACATC CAGCCTATAC TGGCTTATCC TCATCCCAA ACCCTAAAGC AACTAAGGGG 480  
 ATTCCTTGGC GTAATAGGTT TCTGCCGAAA ATGGATTCCC AGGTATGGCG AAATAGCCAG 540  
 GTCATTAAAT AACTAATTA AGGAACTCA GAAAGCCAAT ACCCATTAG TAAGATGGAC 600  
 15 AACTGAAGTA GAAGTGGCTT TCCAGGCCCT AACCAAGCC CCAGTGTTAA GTTGCCAAC 660  
 AGGGCAAGAC TTTGTTCAT ATGTCACAGA AAAACAGGA ATAGCTCTAG GAGTCCTTAC 720  
 20 ACAGATCCGA GGGATGAGCT TGCAACCTGT GGCACACCTG ACTAAGGAAA TTGATGTAGT 780  
 GGCAAAGGGT TGACCTCATT GTTTACGGGT AGTGGTGGCA GTAGCAGTCT TAGTATCTGA 840  
 AGCAGTTAAA ATAATACAGG GAAGAGATCT TACTGTGTGG ACATCTCATG ATGTGAATGG 900  
 25 CATACTCACT GCTAAAGGAG ACTTGTGGCT GTCAGACAAC TGTTTACTTA AATGTCAGGC 960  
 TCTATTACTT GAAGGGCCAG TGCTGCGACT GTGCACTTGT GCAACTCTTA ACCCAGCCAC 1020  
 30 ATTTCTTCCA GACAATGAAG AAAAGATAAA ACATAACTGT CAACAAGTAA TTTCTCAAAC 1080  
 CTATGCCACT CGAGGGGACC TTTTAGAGGT TCCTTTGACT GATCCCGACC TCAACTTGTA 1140  
 TACTGATGGA AGTTCCTTTG TAGAAAAAGG ACTTCGAAAA GTGGGGTATG CAGTGGTCAG 1200  
 35 TGATAATGGA AACTTGAAA GTAATCCCCT CACTCCAGGA ACTAGTGCTC AGCTAGCAGA 1260

ACTAATAGCC CTCACTTGGG CACTAGAATT AGGAGAAGAA AAAAGGGCAA ATATAATACA 1320  
 GACTCTAAAT ATGCTTACCT AGTCCTCCAT GCCCATGCAG CAATATGGAA AGAAAGGGAA 1380  
 5 TTCCTAACTT CTGAGAGAAC ACCTATCAAA CATCAGGAAG CCATTAGGAA ATTATTATTG 1440  
 GCTGTACAGA AACCTAGAGA GGTGGCAGTC TTACTGTC GGGGTCATCA CAAAGGAAAG 1500  
 10 GAAAGGGAAA TACAAGAGAA CTGCCAAGCA TATATTGAAG CCAAAGAGC TGCAAGGCAG 1560  
 GACCCTCCAT TAGAAATGCT TATTAACTT CCCTTAGTAT AGGGTAATCC CTTCCGGGAA 1620  
 ACCAAGCCCC AGTACTCAGC AGGAGAAACA GAATGGGGAA CCTCACGAGG CAGTTTCTC 1680  
 15 CCCTCGGGAC GGTAGCCAC TGAAGAAGGG AAAATACTTT TGCCTGCAAC TATCCAATGG 1740  
 AAATTACTTA AAACCCTTCA TCAAACCTTT CACTTAGGCA TCGATAGCAC CCATCAGATG 1800  
 20 GCCAAATCAT TATTTACTGG ACCAGGCCTT TTCAAACTA TCAAGCAGAT AGTCAGGGCC 1860  
 TGTGAAGTGT GCCAGAGAAA TAATCCCCTG CTTATCGCC AAGCTCCTC AGGAGAACAA 1920  
 AGAACAGGCC ATTACCCTGG AGAAGACTGG CAACTGATTT TACCCACAAG CCCAACCTC 1980  
 25 AGGGATTTC AATCTACTA GTCTGGGTAG ATACTTTCAC GGGTTGGGCA GAGGCCTTCC 2040  
 CCTGTAGGAC AGAAAAGGCC CAAGAGGTAA TAAAGGCACT AGTTCATGAA ATAATCCCA 2100  
 30 GATTCGGACT TCCCCGAGGC TTACAGAGTG ACAATAGCCC TGCTTTCCAG GCCACAGTAA 2160  
 CCCAGGGAGT ATCCAGGCG TTAGGTATAC GATATCACTT AACTGCGCC TGAAGGCCAC 2220  
 AGTCCTCAGG GAAGGTCGAG AAAATGAATG AAACACTCAA AGGACATCTA AAAAAGCAAA 2280  
 35 CCCAGGAAAC CCACCTCACA TGGCCTGTTT TGTGCTTAT AGCCTTAAAA AGAATCTGCA 2340

ACTTTCCCA AAAAGCAGGA CTTAGCCCAT ACGAAATGCT GTATGGAAGG CCCTTCATAA 2400

CCAATGACCT TGTGCTTGAC CCAAGACAGC CAACTTAGTT GCAGACATCA CCTCCTTAGC 2460

5 CAAATATCAA CAAGTTCTTA AAACATTACA AGGAACCTAT CCCTGAGAAG AGGAAAAGAA 2520

TATTCCACCC AAGTGACATG GTATTAGTCA AGTCCCTTCC CTCTAATTCC CCATCCCTAG 2580

10 ATACATCCTG GGAAGGACCC TACCCAGTCA TTTTATCTAC CCCAACTGCG GTTAAAGTGG 2640

CTGGAGTGGA GTCTTGATA CATCACACTT GAGTCAAATC CTGGATACTG CCAAAGGAAC 2700

CTGAAAATCC AGGAGACAAC GCTAGCTATT CCTGTGAACC TCTAGAGGAT TTGCGCCTGC 2760

15 TCTTCAAACA ACAACCAGGA GGAAAATCG AAGCTGTAAA ACTACAAATG GAGCCCAAGA 2820

TGCAGTCCAA GACTAAGATC TACCGCAGAC CCCTGGACCG GCCTGTTAGC CCACGATCTG 2880

20 ATGTTAATGA CATCAAAGGC ACCCCTCCTG AGGAAATCTC AGCTGCACAA CCTCTACTAC 2940

GCCCCAATTC AGCAGGAAGC AGTTAGAGCG GTCGTCGGCC AACCTCCCCA ACAGCACTTA 3000

GGTTTTCTG TTGAGATGGG GGAAGTGAAG ACAGGACTAG CTGGATTTC TAGGCTGATT 3060

25 AAGAATCCCT AAGCCTAGCT GGAAGGTGA CCACATCCAC CTTTAAACAC GGGGCTTGCA 3120

ACTTAGCTCA CACCTGACCA ATCAGAGAGC TCACTAAAAT GCTAATTAGG CAAAGACAGG 3180

30 AGGTAAAGAA ATAGCCAATC ATTTATTGCC TGAGAGCACA GCAGGAGGGA CAATGATCGG 3240

GATATAAACC CAAGTTTTCG AGCCGGCAAC GGCAACCCCC TTTGGGTCCC CTCCCTTTGT 3300

ATGGGAGCTC TGTTTTCATG CTATTTCACT CTATTAAATC TTGCAACTGC AAAAAAAAAA 3360

35 AAAAAAAAAA AA 3372

## (2) INFORMATIONS POUR LA SEQ ID NO:29:

## 5 (i) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 2372 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

## 10 (ii) TYPE DE MOLECULE: ADN.

## (xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:29:

ACTGCACTCT TCTGGTCCAT GTTCTTACG GCTCGAGCTG AGCTTTTGCT CACCGTCCAC 60

CACTGCTGTT TGCCACCACC GCAGACCTGC CGCTGACTCC CATCCCTCTG GATCCTGCAG 120

15 GGTGTCCGCT GTGCTCCTGA TCCAGCGAGG CGCCCATTCG CGCTCCCAAT TGGGCTAAAG 180

GCTTGCCATT GTTCCTGCAC GGCTAAGTGC CTGGGTTTGT TCTAATTGAG CTGAACACTA 240

20 ATCACTGGGT TCCATGGTTC TCTTCTGTGA CCCACGGCTT CTAATAGAAC TATAACACTT 300

ACCACATGGC CCAAGATTCC ATTCCCTTGA ATCCGTGAGG CCAAGAACTC CAGGTCAGAG 360

AATACGAGGC TTGCCACCAT CTTGGAAGCG GCCTGCTACC GTCTTGAAG TGGTTCACCA 420

25 CCATCTTGGG AGCTCTGTGA GCAAGGACCC CCCGGTAACA TTTTGGCAAC CAACGACGGA 480

CATCCAAAGT GATGGGAAAC GTTCCCCGCA AGACAAAAAC GCCCCTAAGA CGTATTCTGG 540

30 AGAATTGGGA CCAATTTGAC CCTCAGACAC TAAGAAAGAA ACGACTTATA TTCTTCTGCA 600

GTGCCGCCTG GCACTCCTGA GGGAAGTATA AATTATAACA CCATCTTACA GCTAGACCTC 660

TTTTGTAGAA AAGGCAAATG GAGTGAAGTG CCATAAGTAC AAACCTTTCTT TTCATTAAGA 720

35 GACAACCTCAC AATTATGTAA AAAGTGTGAT TTATGCCCTA CAGGAAGCCT TCAGAGTCTA 780

CCTCCCTATC CCAGCATCCC CGACTCCTTC CCCAACTAAT AAGGACCCCC CTTCAACCCA 840

AATGGTCCAA AAGGAGATAG ACAAAGGGT AAACAGTGAA CCAAAGAGTG CCAATATTCC 900

5 CCAATTATGA CCCCTCCAAG CAGTGGGAGG AAGAGAATTC GGCCCAGCCA GAGTGCATGT 960

GCCTTTTCT CTCCCAGACT TAAAGCAAAT AAAACAGAC TTAGGTAAAT TCTCAGATAA 1020

10 CCCTGATGGC TATATTGATG TTTTACAAGG GTTAGGACAA TTCTTTGATC TGACATGGAG 1080

AGATATAATG TCACTGCTAA ATCAGACACT AACCCCAAAT GAGAGAAGTG CCACCATAAC 1140

TGCAGCCTGA GGGTTTGGCG TCTCTGGTAT CTCAGTCAGG TCAATGGATA NGGATGACAA 1200

15 CAGAAGGAAA GANAATGATT CCCACAGGC CAGCAGGCAG TTCCCAGTCT AGACCCTCAT 1260

TGGGACACAG AATCAGAACA TGGAGATTGG TGCTGCAGAC ATTTGCTAAC TTGTGTGCTA 1320

20 GAAGGACTAA GGAAACTAG GAAGAAGTCT ATGAATTACT CAATGATGTC CACCATAACA 1380

CAGGGAAGGG AAGAAATCC TACTGCCTTT CTGGAGAGAC TAAGGGAGGC ATTGAGGAAG 1440

CGTGCCTCTC TGTCACCTGA CTCTTCTGAA GGCCAACTAA TCTTAAAGCG TAAGTTTATC 1500

25 ACTCAGTCAG CTGCAGACAT TAGAAAAAC TTCAAAGTC TGCCGTAGGC CCGGAGCAA 1560

ACTTAGAAAC CCTATTGAAC TTGGCAACCT CGGTTTTTTA TAATAGAGAT CAGGAGGAGC 1620

30 AGGCGGAACA GGACAAACGG GATTAAAAA AAGGCCACCG CTTTAGTCAT GACCCTCAGG 1680

CAAGTGGACT TTGGAGGCTC TGGAAAAGG AAAAGCTGGG CAAATTGAAT GCCTAATAGG 1740

GCTTGCTTCC AGTGCGGTCT ACAAGGACAC TTAAAAAAG ATTGTCCAAG TAGAAGTAAG 1800

35 CCGCCCCCTC GTCCATGCCC CTTATTCAA GGAATCACT GGAAGGCCCA CTGCCCCAGG 1860

GGACAAAGGT CTTTGTAGTC AGAAGCCACT AACCAGATGA TCCAGCAGCA GGA CTGAGGG 1920  
 TGCCTGGGGC AAGCGCCATC CCATGCCATC ACCCTCACAG AGCCCTGGGT ATGCTTGACC 1980  
 5 ATTGAGGGCC AGGAAGGTTG TCTCCTGGAC ACTGGTGCGG TCTTCTTAGT CTTACTCTTC 2040  
 TGTCCCGGAC AACTGTCCTC CAGATCTGTC ACTATTCTGA GGGGGTCCNT AAGACGGGCA 2100  
 10 GTCACTAGAT ACTTTTTCCT AGCCACTAAG TTATGAACTG GGGAGCTTTA TTCTTTTCAC 2160  
 ATGCTTTTCT AATTATGCTT GAAAGCCCCA CTACCTTGTT AGGGAGAGAC ATTCTAGCAA 2220  
 AAGCAGGGGC CATTATACAC CTGAACATAG GAGAAGGAAC ACCCGTTTGT TGTNCCCCTG 2280  
 15 CTTGAGGAAG GAATTAATCC TGAAGTCTGG GCAACAGAAG GACAATATGG ACGAGCCAAA 2340  
 GAATGCCCGT CCTGTTCAAG TTAACTAAA GG 2372

20

(2) INFORMATIONS POUR LA SEQ ID NO:30:

(i) CARACTERISTIQUES DE LA SÉQUENCE:

(A) LONGUEUR: 7582 paires de bases

(B) TYPE: nucléotide

25 (C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADN

(xi) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:30:

CAACAATCGG GATATAAACC CAGGCATTCG AGCTGGCAAC AGCAGCCCCC CTTTGGGTCC 60  
 30 CTTCCTTTG TATGGGAGCT GTTTCATGC TATTCACTC TATTAAATCT TGCAACTGCA 120  
 CTCTCTGGT CCATGTTTCT TACGGCTCGA GCTGAGCTTT TGCTCACCGT CCACCACTGC 180  
 TGTTTGCCAC CACCGCANAC CTGCCGCTGA CTCCCATCCC TCTGGATCCT GCAGGGGTGC 240  
 CGCTGTGCTC CTGATCCAGC GARGCGCCCA TTGCCGCTCC CAATTGGGCT AAAGGCTTGC 300  
 CATGTNCCT GCACGGCTAA GTGCCTGGGT TTGTTCTAAT TGAGTGAAC ACTANTCACT 360  
 35 GGGTTCCATG GTTCTCTTCT GTGACCCACG GCTTCTAATA KAACTATAAC ACTTACCACA 420  
 TGGCCCAAGA TTCCATTCCCT TGGAATCCGT GAGGSCAACG AACTCCAGGT CAGAGAATAC 480

GARGCTTGCC ACCATCTTGG AAGCGGCCTG CTACCRCTCTT GGAAGTGGTT CACCACCATC 540  
 TTGGGAGCTC TGTGAGCAAG GACCCCCCGG TRACATTTTG GCRACCAMSR ACGGACATCC 600  
 MAAGTGATGG GAAACGTTCC CCGCAAGACA AAAACGCCCC TAAGACGTAT TCTGGARAAT 660  
 TGGGAMCAAT TTGACCCTCA GACACTAAGA AAGAAACGAC TTATATTCTT CTGCAGTGCC 720  
 5 GCCTGGCACT CCTGAGGGAA GTATAAATTA TAACACCATC TTACAGCTAG ACYTCTTTTG 780  
 TAGAAAAGGC AAATGGAGTG AAGTGCCATA AGTACAACT TTCTTTTCAT TAAGAGACAA 840  
 CTCACAATTA TGTAAAAAGT GTGATTATG CCCTACAGGA AGCCTTCAGA GTCTACCTCC 900  
 CTATCCCAGC ATCCCCGACT CCTTCCCCAM YTAATAAGGA CCCCCCTTCA ACCCAAATGG 960  
 TCCAAAAGGA GATAGACAAA AGGGTAAACA GTGAACCAA GAGTGCCAAT ATTCCCCAAT 1020  
 10 TATGACCCCT CCCAAGCAGT GGGAGGAAGA GAATTCGGCC CAGCCAGAGT GCATGTGCYT 1080  
 TTTYTCTCC CAGACTTAAA GCAAATAAAA ACAGACTTAG GTAAATTCTC AGATAAYCCT 1140  
 GATGGCTATA TTGRTGTTTT ACAAGGGTTA GGACAATTCT TTGATCTGAC ATGGAGAGAT 1200  
 ATATATGTCA CTGCTAAATC AGACACTAAC CCCAAATGAG AGAAGTGCCA CCATAACTGC 1260  
 AGCCTGAGRG TTTGGCGATC TCTGGTATCT CAGTCAGGTC AATGGATANG GATGACAACA 1320  
 15 GAAGGAAAGA NAATGATTCC CCACAGGCCA GCARGCAGTT CCCAGTCTAS ACCCTCATTTG 1380  
 GGGACACAGA AATCAGTAAC ATGGGAGATT GGTGCTGCAG ACATTTGCTA ACTTGTGTGC 1440  
 TASAAGGACT AAGGAAAAC ASGAAGAAAR TCTAYGAATT ACTCAATGAT GTCCACCATA 1500  
 ACACAGGGGA AGGGAAGAAA ATCCTACTGC CTTTCTGGAG AGACTAAGGG AGGCATTGAG 1560  
 GAAGCGTGCC TCTCTGTAC CTGACTCTTC TGAAGGCCAA CTAATCTTAA AGCGTAAGTT 1620  
 20 TATCACTCAG TCAGCTGCAG ACATTAGAAA AAACCTCAAA AGTCTGCCGT AGGCCCGGAG 1680  
 CAAACTTAG AAACCTTATT GAACTTGGCA ACYTCTGTTT TTTATAATAG AGATCAGGAG 1740  
 GAGCAGGCGG AACAGGACAA ACGGGATTAA AAAAAAGGCC ACCGCTTTAG TCATGACCCT 1800  
 CAGGCAAGTG GACTTTGGAG GCTCTGGAAG AGGGAAAAGC TGGGCAAATT GAATGCCTAA 1860  
 TAGGGCTTGC TTCCAGTGCG GTCTACAAGG ACACTTTAAA AAAGATTGTC CAAGTAGAAG 1920  
 25 TAAGCCGCCC CTTCTGCCAT GCCCCTTATT TCAAGGGAAT CACTGGAAGG CCCACTGCCC 1980  
 CAGGGGACAA AGGTCTTTTG AGTCAGAAGC CACTAACCAG ATGATCCAGC AGCAGGACTG 2040  
 AGGGTGCCCTG GGGCAAGCGC CATCCCATGC CATCACCTC ACAGAGCCCT GGGTATGCTT 2100  
 GACCATTGAG GGCCAGGAAG GTTGTCTCCT GGACACTGGT GCGGTCTTCT TAGTCTTACT 2160  
 CTTCTGTCCC GGACAACTGT CCTCCAGATC TGTCACTATT CTGAGGGGGT CCNTAAGACG 2220  
 30 GGCAGTCACT AGATACTTTY TCCAGCCAC TAAGTTATGA ACTGGGGAGC TTTATTCTTT 2280  
 TCACATGCTT TTCTAATTAT GCTTGAAAGC CCCACTACCT TGTAGGGAG AGACATTCTA 2340  
 GCAAAAGCAG GGGCCATTAT ACACCTGAAC ATAGGAGAAG GAACACCCGT TTGTGTNCC 2400  
 CCTGCTTGAG GAAGGAATTA ATCCTGAAGT CTGGGCAACA GAAGGACAAT ATGGACGAGC 2460  
 CAAAGAATGC CCGTCTGTT CAAGTTAAAC TAAAGGATTC CACTTCCTTT CCCTACCAA 2520  
 35 GGCAGTACCC CCTCAGACCC AAGGCCAAC AAGGATTCCA AAAGATTGTT AAGGACTTAA 2580  
 AAGCCCAAGG CTTAGTAAAA CCATGCATAA CTCCTGCAG TAATTCCGTA GTGGATTGAG 2640

GAGGCACAGA AACCCAGTGG ACAGTGGAGG GTTAGTGCAA GATCTCAGGA TTATCAATGG 2700  
 AGGCCGTTGT CCTTTTATAC CCAGCTGTAC CTAGCCCTTA TACTGTGMYT TCCCAAATAC 2760  
 CAGAGGAAGC AGAGTGGTTT ACASTCCTGG ACCTTMAGGA TGCCTTCTTC TGCATCCCTG 2820  
 TACATCCTGA CTCTCAATTC TTGTTTGCTT TTGAAGATAC TTCAAACCCA RCATCTCAAC 2880  
 5 TCACCTGGAC TRTTTTACCC CAAGGGTTCA GGGATAGYCC CCATCTATTT GGCCAGGCAT 2940  
 TAGCCCAAGA CTTGAGYCAR TYMTCATACC TGGACACTCT TGTCTTCRG TAKGTGGATG 3000  
 ATTTACTTTT RGCYGCCYRT TCAGAAACCT TGTGCCATCA AGCCACCCAA GCRCTCTTMA 3060  
 ATTTCTCTCG YACCTGTGGC TACAWGGTTT CCAAACSARA RGCTCARCTC TGCTCACAGC 3120  
 AGGTTAAATA CTTAGGRCTA ARATTATCCA AAGGCACCAR GGCCCTCAGT GAGGAAYRYA 3180  
 10 TCCAGCCTAT ACTGGCTTAT CCTCATCYCA AAACCCTAAA GCAACTAAGR GRRTTCCTTG 3240  
 GCRTAAYAGG YTTCTGCCGA AWATGGATTC CCCAGGTWTG GCRAATAGC CAGGYCATT 3300  
 WATACASTAA TTAAGGAAAC TCAGAAAGCC AATACCCATT TARTAAGATG GAYAMCTGAA 3360  
 GYMRAAGTGG CTTTCCAGGC CCCTAAAGAA GGCCTTAAAC CCAAGYCCCA GTGTTAAGYT 3420  
 TGCCAACRGG GCAAGACTTT TSTTYATAYR TCACAGAAAA AACAGRAAY AGCTCTRGGA 3480  
 15 GTCCTTACAC AGRTCCRAGG GAYGAGCTTG CAACCYRTGG CRYACCTGAS TAAGGAAAYT 3540  
 GATGTAGTGG CAAAGGGTTG RCYTCATTGT TTAYGGGTAG TGGTGGCAGT AGCAGTYKTA 3600  
 GTATCTGAAG CAGTTAAAAT AATACAGGGR AGAGATCTTA CTGTGTGGAC ATCTCATGAK 3660  
 GTGAAYRGCA TACTCACTGC TAAAGGAGAC TTGTGGCTGT CAGACAACYG TTTACTTAAA 3720  
 TRTCAGGCTC TATTACTTGA ARGGCCAGTG CTGCRCTGT GCACTTGTGC AACTCTTAAC 3780  
 20 CCAGYCNCAT TTCTCCAGA CAATGAAGAA AAGATARAAY ATAACGTCA ACAARTATT 3840  
 TCTCAAACCT ATGCCACTCG AGGGGACCTT YTAGARGTTC CYTTGACTGA TCCYGACCTT 3900  
 CAACTTGTAT ACTGATGGAA GTTCCTTTGT AGAAAAAGGA CTTCGAAAAG YGGGGTATGC 3960  
 AGTGGTCAGT GATAATGGAA TAYTTGAAAG TAATCCCCTC ACTCCAGGAA CTAGTGCTYA 4020  
 GCTRGCAGAA CTAATAGCCY TCAYTKGGGC ACTAGAATTA GGAGAAGRAA AAAGGGYAAA 4080  
 25 TATATATACA GACTCTRART ATGCTYACCT AGTCNTCCAT GCCCATGMRG CAATATGSAR 4140  
 AGAAAGGGAA TTCCTAATT CYGAGRGAC ACCTATCAMA CATCAGGAAG CCATTAGGAR 4200  
 ATTATTAYTG GCWGTACAGA AACCTARAGA GGTGGMAGTC TTACTGTCY GGGGTCATCA 4260  
 NAAAGGAAAG RAAAGGGAAA TASAAGRGAA YTGCCAAGCA KATATTGAAG CMAAAAGAGC 4320  
 TGCAAGGCAG GACCCTCCAT TAGAAATGCT TATTAACTT CCCTTAGTAT AGGGTAATCC 4380  
 30 CTTCGGGGAA ACCAAGCCCC AGTACTCAGC AGGAGAAACA GAATGGGGAA CCTCACGAGG 4440  
 CAGTTTCTC CCCTCGGGAC GGTTAGCCAC TGAAGAAGGG AAAATACTTT TGCCTGCAAC 4500  
 TATCCAATGG AAATTACTTA AAACCCTTCA TCAAACCTT CACTTAGGCA TCGATAGCAC 4560  
 CCATCARATG GCCAAATCAT TATTTACTGG ACCAGGCCTT TTCAAACTA TCAAGCARAT 4620  
 AKTCAGGGCC TGTGAATGT GCCARARAAA TAATCCCCTG CCTYATCGCC AAGCTCCTTC 4680  
 35 AGGARAACAA ARAACAGGCC ATTACCCTGR ARAARACTGG CAACTGATTT TACCCACAAG 4740  
 CCCAAACCTC AGGGATTTCAT GTATCTACTA GTCTGGGTAR ATACTTTCAC GGGTTGGGCA 4800

RAGGCCTTCC CCTGTAGGAC AGAAAAGGCC CAAGAGGTAA TAAAGGCACT AGTTCATGAA 4860  
 ATAATTCCCA GATTCCGACT TCCCCGAGGC TTACAGAGTG ACAATAGCCC TGCTTTCCAG 4920  
 GCCACAGTAA CCCAGGGAGT ATCCCAGGCG TTAGGTATAC GATATCACTT AACTGCGCC 4980  
 TGAAGGCCAC AGTCCTCAGG GAAGGTCGAG AAAATGAATG AAAYACTCAA AGGACATCTA 5040  
 5 AAAAGCAAA CCCAGGAAAC CCACCTCACA TGGCCTGYTC TGTTGCCTAT AGCCTTAAAA 5100  
 AGAATCTGCA ACTTTCCCA AAAAGCAGGA CTTAGCCCAT ACGAAATGCT GTATGGAAGG 5160  
 CCCTTCATAA CCAATGACCT TGTGCTTGAC CCAAGACAGC CAACTTAGTT GCAGACATCA 5220  
 CCTCCTTAGC CAAATATCAA CAAGTTCTTA AAACATTACA AGGAACCTAT CCCTGAGAAG 5280  
 AGGGAAGA ACTATTCCAC CCWWTGACA TGGTATTAGT CAAGTCCCTT CYCTCTAATT 5340  
 10 CCCCATCCCT AGATACATCC TGGGAAGGAC CCTACCCAGT CATTTTATYT ACCCCAAGT 5400  
 CGGTTAAGT GGCTGGAGTG GAGTCTTGA TACATCACAC TTGAGTCAA TCCTGGATAC 5460  
 TGCCAAAGGA ACCTGAAAAT CCAGGAGACA ACGCTAGCTA TTCTGTGAA CCTCTAGAGG 5520  
 ATTTGCGCCT GCTCTTCAA CAACAACCAG GAGGAAAGTA ACTAAAATCA TAAATCCCCC 5580  
 ATGGSCTCC CTTATCATAT TTTTCTCTKT ASTGTTSTTT YACCCTSTTT CACTCTCACT 5640  
 15 GCACCCCTC CATGCCGCTG TATGACCACT AGCTCCCCTY ACCMAGAGTT TCTATGGAGA 5700  
 ATGCAGCGTC CCGGAAATAT TGATGCCCCA TCGTATAGGAG TCTTTSTAAG GGAACCCCC 5760  
 ACCTTCACTG CCCACACCCA TATGCCCCGC AACTGCTATC ACTCTGCCAC TCTTTGCATG 5820  
 CATGCAAATA CTCATTATTG GACAGGAAAA ATGATTAATC CTAGTTGTCC TGGAGGACTT 5880  
 GGAGTCACTG TCTGTTGGAC TTAATTCACC CAACTGGTA TGTCTGATGG GGGTGGAGTT 5940  
 20 CAAGATCAGG CAAGAGAAAA ACATGTAAAA GAAGTAATCT CCAACTCAC CSGGGTACAT 6000  
 GGCACCTCTA GCCCCTACAA AGGACTAGAT CTCTCAAAC TACATGAAAC CCTCCGTACC 6060  
 CATACTCGCC TGGTAAGCCT ATTTAATACC ACCCTCACTG GGCTCCATGA GGTCTCGGCC 6120  
 CAAAACCTA CTAATGTTG GATATGCCTC CCCCTGAACT TCARGCCATA TGTTTCAATC 6180  
 CCTGTACCTG AACAATGGAA CAACTTCAGC ACAGAAATAA ACACCACTTC CGTTTTAGTA 6240  
 25 GGACCTCTTG TTTCCAATST GGAAATAACC CATACCTCAA ACCTCACCTG TGTAATAATT 6300  
 AGCAATACTA CATACACAAC CAACTCCCAA TGCATCAGGT GGGTAACTCC TCCCACACAA 6360  
 ATAGTCTGCC TACCCTCAGG AATATTTTT GTCTGTGGTA CCTCAGCCTA TCGTTGTTG 6420  
 AATGGCTCTT CAGAATCTAT GTGCTTCCTC TCATTCTTAG TGCCCCCYAT GRCCATCTAC 6480  
 ACTGAACAAG ATTTATACAG TTATGTCATA TCTAAGCCCC GCAACAAAAG AGTACCCATT 6540  
 30 CTTCTTTTG TTATAGGAGC AGGAGTGCTA GGTGCACTAG GTACTGGCAT TGGCGGTATC 6600  
 ACAACCTCTA CTCAGTTCTA CTACAACTA TCTCAAGAAC TAAATGGGGA CATGGAACGG 6660  
 GTCGCCGACT CCCTGGTCAC CTTGCAAGAT CAACTTAACT CCCTAGCAGC AGTAGTCTT 6720  
 CRAAATCGAA GAGCTTTAGA CTYGCTAACC GCTGARAGAG GGGGAACCTG TTTATTTT 6780  
 GGGGAAGAAT GCTGTTATTA TGTTAATCAA TCCGGAATCG TCACTGAGAA AGTTRAAGAA 6840  
 35 ATTCAGATC GAATACAACG TAKAGCAGAR GAGCTTCGAA AACTGGACC CTGGGGCTC 6900  
 CTCAGCCRAT GGATGCCCTG GATTCTCCCC TTCTTAGGAC CTCTAGCAGC TATAATATTG 6960

CTACTCCTCT TTGGACCCCTG TATCTTTTRAC CTCCTTGTTA ACTTTGTCTC TTCCAGAATC 7020  
 GAAGCTGTRA AACTACAAAT GGAGCCCAAG ATGCAGTCCA AGACTAAGAT CTACCGCAGA 7080  
 CCCCTGGACC GGCCTGYTAG CCCACGATCT GATGTTAATG ACATCAAAGG CACCCCTCCT 7140  
 GAGGAAATCT CAGCTGCACA ACCTCTACTA CGCCCCAATT CAGCAGGAAG CAGTTAGAGC 7200  
 5 GGTSGTCGGC CAACCTCCCC AACAGCACTT AGGTTTTCCT GTTGAGATGG GGGACTGAGA 7260  
 GACAGGACTA GCTGGATTTC CTAGGCTGAY TAAGAATCCY TAAGCCTAGS TGGGAAGGTG 7320  
 ACCACATCCA CCTTTAAACA CGGGGCTTGC AACTTAGYTC ACACCTGACC AATCAGAGAG 7380  
 CTCATAAAA TGCTAATTAG GCAAAGACAG GAGGTAAAGA AATAGCCAAT CATYTATTGC 7440  
 MTGAGAGCAC AGCAGGAGGG ACAATGATCG GGATATAAAC CCAAGTYTTC GAGCCGGCAA 7500  
 10 CGGCAACCCC CTTTGGGTCC CCTCCCTTTC TATGGGAGCT CTGTTTTTCAT GCTATTTTCAC 7560  
 TCTATTAAAT CTTGCARCTG CR 7582

## (2) INFORMATIONS POUR LA SEQ ID NO:31:

## (i) CARACTERISTIQUES DE LA SÉQUENCE:

- 15 (A) LONGUEUR: 363 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADN

## 20 (x1) DESCRIPTION DE LA SÉQUENCE: SEQ ID NO:31:

MGNIPPKAKT	PLRCILENWD	QCDTQTLRKK	RFIFPCSTAW	PQYPLQGRET	50
WLPEGSINYN	IIQLDLFCR	KEGKWSEVPY	VQTFPSLRDN	SQLCKKGLC	100
PTGSPQSPFP	YPSVPPPTPS	STNKDPPLTQ	TVQKEIDKGV	NNEPKSANIP	150
RLCPLQAVRG	GEFGPARVPV	PFSLSDLKQI	KIDLKGFSDN	PDGYIDVLQG	200
25 LGQSFDLTWR	DIIMLLNQTL	TPNERSAAVT	AAREFGDLWY	LSQVNNRMTT	250
EERTTPTGQQ	AVPSVDPHWD	TESEHGDWCH	KHLITCVLEG	LRKTRKKPMN	300
YSMMSTITQG	KEENPTAFLD	RLREALRKHT	SLSPDSIEGQ	LILKDKFITQ	350
SAADIRKNFK	SLP				363

30

## REVENDEICATIONS

1. Fragment nucléaire endogène, de type rétroviral, à l'état isolé, et intégré à l'ADN du génome humain, ledit fragment étant caractérisé en ce qu'il  
5 comprend, ou consiste en, au moins une partie du gène gag d'un rétrovirus endogène associé à une maladie auto-immune ou à des insuccès de grossesse ou pathologies de la grossesse, ladite partie au moins codant, directement ou indirectement, pour un produit d'expression, ou le  
10 complémentaire dudit fragment.
2. Fragment selon la revendication 1, caractérisé en ce qu'il comprend, ou consiste en, ledit gène gag entier.
3. Fragment selon la revendication 1, caractérisé  
15 en ce que la partie au moins code pour la matrice et la capsid.
4. Fragment selon la revendication 1, caractérisé en ce qu'il comprend SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, ou le complémentaire de l'une quelconque  
20 desdites séquences.
5. Fragment selon la revendication 1, caractérisé en ce qu'il consiste en SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, ou le complémentaire de l'une quelconque desdites séquences.
- 25 6. Fragment selon la revendication 1, caractérisé en ce qu'il est localisé sur au moins l'un des chromosomes humain 1, 3, 6, 7 et 16.
7. Fragment selon la revendication 6, caractérisé en ce qu'il est localisé sur au moins le chromosome 3.
- 30 8. Fragment selon la revendication 1, caractérisé en ce que le produit d'expression est de l'ARN messager.
9. Fragment selon la revendication 1, caractérisé en ce que le produit d'expression est immunologiquement reconnu par des anticorps présents dans un échantillon  
35 biologique d'un patient atteint d'une maladie auto-immune.

10. Fragment selon la revendication 9, caractérisé en ce que l'échantillon biologique est un fluide biologique choisi parmi le sérum, le plasma, le liquide synovial et l'urine.

5 11. Fragment selon la revendication 9, caractérisé en ce que la maladie auto-immune est la sclérose en plaques.

12. Produit de transcription endogène, à l'état isolé, susceptible d'être obtenu par transcription d'au moins  
10 ladite partie du gène gag d'un fragment selon l'une quelconque des revendications 1 à 11.

13. Procédé de détection de séquences nucléotidiques endogènes appartenant à un fragment tel que défini selon l'une des revendications 1 à 11, caractérisé en ce que :

15 - on effectue au préalable une étape d'extraction de l'ADN cellulaire d'un tissu ou fluide biologique, on réalise au moins un cycle d'amplification de l'ADN cellulaire, telle que par PCR, au moyen d'amorces notamment choisies parmi SEQ ID NO:4 à SEQ ID NO:9 et  
20 SEQ ID NO :12 à SEQ ID NO :17,

- on met en contact l'ADN cellulaire présent dans l'échantillon avec une sonde déterminée susceptible de s'hybrider avec un fragment tel que défini précédemment et de former un complexe d'hybridation, ladite sonde  
25 comprenant au moins 15 nucléotides, de préférence 17 et avantageusement 19 nucléotides contigus de SEQ ID NO :3 ou consistant en SEQ ID NO :3, dans des conditions appropriées permettant l'hybridation, en particulier dans des conditions de forte stringence, et

30 - on détecte les complexes d'hybridation formés par tout moyen approprié.

14. Procédé selon la revendication 13, caractérisé en ce que la sonde est marquée par un traceur, tel que par exemple un traceur radioactif ou une enzyme.

15. Procédé de détection de séquences nucléotidiques endogènes appartenant à un fragment tel que défini selon l'une des revendications 1 à 11, caractérisé en ce que :

- 5 - on effectue au préalable une étape d'extraction de l'ADN cellulaire d'un tissu ou fluide biologique, éventuellement issu de chromosomes isolés, on réalise au moins un cycle d'amplification de l'ADN cellulaire, telle que par PCR, au moyen d'amorces notamment choisies parmi SEQ ID NO:4 à  
10 SEQ ID NO:9 et SEQ ID NO :12 à SEQ ID NO :17,
  - on effectue une étape de transcription / traduction in vitro du produit amplifié, et
  - on fait réagir le produit issu de l'étape de transcription / traduction avec un sérum ou plasma d'un  
15 patient présentant une maladie auto-immune.

16. Procédé pour l'étude et/ou le suivi d'une prolifération cellulaire des cellules T in vitro selon lequel on met en contact les cellules T d'un patient avec soit des produits de transcription / traduction (SEQ ID  
20 NO :31), tels qu'obtenus selon le procédé de la revendication 15, soit des peptides de synthèse dérivés de ou appartenant à SEQ ID NO :31.

17. Procédé de marquage moléculaire in situ de chromosomes isolés de patients dans lequel on utilise une  
25 sonde marquée, par tout traceur approprié, comprenant tout ou partie de SEQ ID NO :3.

18. Protéine recombinante obtenue à partir d'une cassette d'expression dans un hôte bactérien caractérisée en ce que sa séquence protéique consiste en SEQ ID NO :31.

30 19. Protéine, selon la revendication 18, caractérisée en ce que l'hôte bactérien est *E. coli*.

20. Réactif pour la détection d'une maladie auto-immune ou un suivi de grossesse comprenant au moins un fragment selon l'une quelconque des revendications 1 à 11  
35 ou une protéine selon les revendications 16, 18 et 19.

21. Utilisation d'un fragment selon l'une quelconque des revendications 1 à 11 ou d'une protéine selon les revendications 16, 18 et 19 pour détecter, dans un échantillon biologique, une susceptibilité à une  
5 maladie auto-immune ou un suivi de grossesse.

22. Utilisation selon la revendication 21, caractérisée en ce que la maladie auto-immune est la sclérose en plaques.

## REVENDEICATIONS

1. Fragment nucléaire endogène, de type rétroviral, à l'état isolé, ledit fragment étant caractérisé en ce qu'il comprend, ou consiste en, au moins une partie du  
5 gène gag d'un rétrovirus endogène associé à une maladie auto-immune ou à des insuccès de grossesse ou pathologies de la grossesse, ladite partie au moins codant, directement ou indirectement, pour un produit d'expression, ou le complémentaire dudit fragment.
- 10 2. Fragment selon la revendication 1, caractérisé en ce qu'il comprend, ou consiste en, ledit gène gag entier.
3. Fragment selon la revendication 1, caractérisé en ce que la partie au moins code pour la matrice et la  
15 capsid.
4. Fragment selon la revendication 1, caractérisé en ce qu'il comprend SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, ou le complémentaire de l'une quelconque desdites séquences.
- 20 5. Fragment selon la revendication 1, caractérisé en ce qu'il consiste en SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, ou le complémentaire de l'une quelconque desdites séquences.
6. Fragment selon la revendication 1, caractérisé  
25 en ce qu'il est localisé sur au moins l'un des chromosomes humain 1, 3, 6, 7 et 16.
7. Fragment selon la revendication 6, caractérisé en ce qu'il est localisé sur au moins le chromosome 3.
8. Fragment selon la revendication 1, caractérisé  
30 en ce que le produit d'expression est de l'ARN messager.
9. Fragment selon la revendication 1, caractérisé en ce que le produit d'expression est immunologiquement reconnu par des anticorps présents dans un échantillon biologique d'un patient atteint d'une maladie auto-immune.

FIG 1

	Noms	Similitudes	Répétitions	ORFs
28274	ARN Recons	96%	oui	538
6911	RG083M05 [7]	88%	oui	538
35199	BAC378 [14]	89%	oui	non
91299	Q11M15 [21]	88%	oui	413 et 305
	U134E6 [x]		non	non

FIG 2

	10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890		
OCTAGAACGT	ATTCTGGAGA	ATTGGGACCA	ATGTGACACT	CAGACGCTAA		50
P R T Y	S G E	L G P	M .	H S D A K		
L E R	I L E N	W D Q	C D T	Q T L R		
. N V	F W R	I G T N	V T L	R R .		
GAAAGAAAG	ATTATATTC	TCTGCGAGTA	CGGCTGGCC	ACAATATCT		100
K E T	I Y I L	L Q Y	R L A	T I S S		
K K R	F I F	F C S T	A W P	Q Y P		
E R N D	L Y S	S A V	P P G H	N I L		
CTTCAAGGGA	GAGAACTG	GCTTCTGAG	GGAAGTATAA	ATTATAACAT		150
S R E	R N L	A S .	G K Y K	L . H		
L Q G R	E T W	L P E	G S I N	Y N I		
F K G	E K P G	F L R	E V .	I I T S		
CATCTTACAG	CTAGACCTCT	TCTGTAGAAA	GGAGGGCAAA	TGGAGTGAAG		200
H L T A	R P L L	. K G G	Q M E .	S		
I L Q	L D L F	C R K	E G K	W S E V		
S Y S .	T S	S V E R	R A N	G V K		
TGCCATATGT	GCAACTTTC	TTTTCATTAA	GAGACAATC	ACAATTATGT		250
A I C	A N F L	F I K	R Q L	T I M .		
P Y V	Q T F	F S L R	D N S	Q L C		
C H M C	K L S	F H .	E T T H	N Y V		
AAAAAGTGTG	GTTTATGCCC	TACAGGAAGC	OCTCAGAGTC	CACTTCCCTA		300
K V W	F M P	Y R K P	S E S	T S L		
K K C G	L C P	T G S	P Q S P	P P Y		
K S V	V Y A L	Q E A	L R V	H L P T		
CCCCAGGTC	CCCTCCCCGA	CTCCTTCTC	AACTAATAAG	GACCCCCCTT		350
P Q R P	L P D	S F L N .	. G P P F			
P S V	P S P T	P S S	T N K	D P P L		
P A S	P P R	L L P Q	L I R	T P L		
TATCCCAAC	GGTCCAAAG	GAGATAGACA	AAGGGGTAAA	CAATGAACCA		400
N P N	G P K G	D R Q	R G K	Q . T K		
T Q T	V Q K	E I D K	G V N	N E P		
. P K R	S K R R .	T K G .	T M N Q			
AAGAGTGCCA	ATATTCCCCG	ATTATGCCCC	CTCCAAGCAG	TCAGAGGAGG		450
E C Q	Y S P	I M P P	P S S	E R R		
K S A N	I P R	L C P	L Q A V	R G G		
R V P	I F P D	Y A P	S K Q .	. E E E		
AGANTTGGC	CCAGGCAGAG	TGCTGTAC	TTTTTCTCTC	TCAGACTTAA		500
R I R P	S Q S	A C T	F F S L	R L K		
E F G	P A R V	P V P	F S L	S D L K		
N S A	Q P E	C L Y L	F L S	Q T .		

FIG 2 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AGCAAATTAA	AATAGACCTA	GGTAAATTCT	CAGATAAACC	TGACGGCTAT	550
A N .	N R P R .	I L R .	P .	R L Y	
Q I K	I D L	G K F S	D N P	D G Y	
S K L K .	T .	V N S	Q I T L	T A I	
ATTGATGTTT	TACAAGGGTT	AGGACAATOC	TTTGATCTGA	CATGGAGAGA	600
. C F	T R V	R T I L .	S D M E R		
I D V L	Q G L	G Q S	F D L T	W R D	
L M F	Y K G .	D N P	L I .	H G E I	
TATAATGTTA	CTACTAAATC	AGACACTAAC	CCCAATGAG	AGAAGTGOOG	650
Y N V T	T K S	D T N	P K .	E K C R	
I M L	L L N Q	T L T	P N E	R S A A	
. C Y Y .	I R H .	P	Q M R	E V P	
CTGTAACTGC	AGCCCGAGAG	TTTGGCGATC	TTTGGTATCT	CAGTCAGGOC	700
C N C	S P R V	W R S	L V S	Q S G Q	
V T A	A R E	F G D L	W Y L	S Q A	
L .	L Q	P E S	L A I	F G I S , V R P	
AACAATAGGA	TGACAACAGA	GGAAAGAACA	ACTCCACAG	GCCAGCAGGC	750
Q .	D D N R	G K N N	S H R	P A G	
N N R M	T T E	E R T	T P T G	Q Q A	
T I G .	Q Q R	K E Q	L P Q	A S R Q	
AGTTCOCAGT	GTAGACCCTC	ATTGGGACAC	AGAATCAGAA	CATGGAGATT	800
S S Q C	R P S	L G H	R I R T	W R L	
V P S	V D P H	W D T	E S E	H G D W	
F P V .	T L	I G T Q	N Q N	M E I	
GGTGGCAGAA	ACATTGCTA	ACTTGGGTGC	TAGAAGGACT	GAGGAAACT	850
V P Q	T F A N	L R A	R R T	E E N .	
C H K	H L L	T C V L	E G L	R K T	
G A T N	I C .	L A C .	K D .	G K L	
AGGAAGAACC	CTATGAATTA	CTCAATGATG	TOCACTATAA	CACAGGGAAA	900
E E A	Y E L	L N D V	H Y N	T G K	
R K K P	M N Y	S M M	S T I T	Q G K	
G R S	L .	I T Q .	C P L .	H R E R	
GGAGAGAAAT	CTTACTGCTT	TTCITGGACAG	ACTAAGGGAG	GCATTGAGGA	950
G R K S	Y C F	S G Q	T K G G	I E E	
E E N	L T A F	L D R	L R E	A L R K	
K K I	L L L	F W T D .	G R H .	G	
AGCATACCTC	CCTGTCACTT	GACTCTATTG	AAGGCCAACT	AATCTTAAAG	1000
A Y L	P V T .	L Y .	R P T	N L K G	
H T S	L S P	D S I E	G Q L	I L K	
S I P P	C H L	T L L	K A N .	S . R	

FIG 2 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GATAAGTTTA	TCACTCAGTC	AGCTGCAGAC	ATTAGAAAAA	ACTTCAAAAG	1050
. V Y H S V	S C R H .	K K L Q K			
D K F I T Q S	A A D I R K N	F K S			
I S L S L S Q	L Q T L E K T	S K V			
TCTGCTTAG <sup>1</sup>	GGGCGAGCA	GAACCTAGAA	ACCTATTTA	ACTTGGCATC	1100
S A L G P E Q	N L E T L F N	L A S			
L P . A R S R	T . K P Y L T	W H P			
C L R P G A	E L R N P I .	L G I			
CTCAGTTTTT	TATAATAGAG	ATCAGGAGGA	GCAGGCGAAA	CGGACAAAC	1150
S V F Y N R D	Q E E Q A K R	D K R			
Q F F I I E	I R R S R R N	G T N			
L S F L . .	R S G G A G E T	G Q T			
GGGATAAAAA	AAAAAGGGG	GGTCCACTAC	TTTAGTCATG	GGCTCAGGC	1200
D K K K R G	G P L L .	S W P S G			
G I K K K G G	V H Y F S H G	P Q A			
G . K K K G G	S T T L V M	A L R Q			
AAGCAGACTT	TGGAGGCTCT	GCAAAAGGGA	AAAGCTGGGC	AAATCAAATG	1250
K Q T L E A L	Q K G K A G Q	I K C			
S R L W R L C	K R E K L G	K S N A			
A D F G G S	A K G K S W A	N Q M			
OCTAATAGGG	CTGGCTTCCA	GTCGGTCTA	CAAGGACACT	TTAAAAAGA	1300
L I G L A S S	A V Y K D T	L K K I			
. . G W L P	V R S T R T L .	K R			
P N R A G F Q	C G L Q G H F	K K D			
TTATCCAAGT	AGAAATAGC	GGGGGCTTG	TCCATGCCCC	TTAGTCAAG	1350
I Q V E I S	R P L V H A P	Y V K			
L S K . K .	A A P L S M P L	T S R			
Y P S R N K P	P P C P C P	L R Q G			
CGAATCACTG	GAAGGCCCAC	TGCCCCAGGG	GATCAAGATA	CCTCAGTCA	1400
G I T G R P T	A P G D E D T	L S Q			
E S L E G P L	P Q G M K I L .	V R			
N H W K A H	C P R G .	R Y S E S			
GAAGCATTTA	ACCAGATGAT	CCAGCAGCAG	GACTGAGGGT	GGGCGGGGG	1450
K P L T R .	S S S R T E G	A R G E			
S H . P D D	P A A G L R V	P G A			
E A I N Q M I	Q Q Q D .	G C P G R			
AGGCGCAGCC	CATGCCATCA	CCCTCAGAGA	GGGCGGGTGA	TGTTTGACCA	1500
R Q P M P S	P S Q S P G Y	V . P			
S A S P C H H	P H R A P G M	F D H			
A P A H A I T	L T E P R V	C L T I			

5/5

FIG 2 (note)

-----

	10	20	30	40	50
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
TTGACAGCCA A					1511
L R A					
. E P					
E S Q					

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☒ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**